

Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio

Práctica Profesional - Licenciatura en Administración



Alumna: Passini, Rosario Lourdes
Tutor: Medina Galván, Marcelo Enrique



INDICE

DATOS GENERALES	1
INFORME FINAL DEL TRABAJO DE PRACTICA PROFESIONAL.....	1
CUMPLIMIENTO DEL PLAN DE TRABAJO ORIGINAL	1
DIVULGACIÓN	2
CURSOS Y/O ESTADÍAS DE CAPACITACIÓN.....	6
BALANCE DE LA EXPERIENCIA EN LA PRÁCTICA PROFESIONAL.....	8
COMENTARIOS DEL DIRECTOR	9
DOCUMENTACIÓN PROBATORIA.....	10
PLAN DE TRABAJO.....	11
INFORMES DE AVANCE.....	22
TRABAJO FINAL	26
ANEXOS.....	55



DATOS GENERALES

Apellido y Nombres : Passini, Rosario Lourdes	Apellido y Nombres del Director: Medina Galván, Marcelo Enrique
DNI 38.487.401	DNI 25.922.471
Tema: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: Aplicaciones y Casos de Estudio	
Opción de Práctica Profesional: Proyecto de Investigación	
Lugar de Trabajo: Facultad de Ciencias Económicas - UNT	

INFORME FINAL DEL TRABAJO DE PRACTICA PROFESIONAL

Las tareas realizadas vinculadas con mi práctica profesional fueron las siguientes: Lectura de dos libros de análisis multivariante, un libro de analítica de negocios, artículos sobre big data, minería de datos, data science, análisis factorial y funcionamiento del software RStudio. Luego de haber comprendido el tema relacionado a mi PP, procedí a buscar las bases de datos con las cuales iba a trabajar, filtré las variables necesarias para el análisis, cargué la matriz en el software y apliqué un análisis factorial, grafiqué los datos e interpreté los resultados para cada una de las bases de datos tanto en forma individual y en conjunto para poder realizar un análisis global. Además, me reuní con mi tutor para controlar el grado de avance de mi trabajo y los resultados parciales que fui obteniendo. Participé de una Reunión de Discusión de Avances de la PP, reuniones de tutoría y de la VI Muestra Académica de Trabajos de Investigación de LAD.

CUMPLIMIENTO DEL PLAN DE TRABAJO ORIGINAL

100%	75%	50%	25%	menos del 25%
	X			

El grado de cumplimiento de la PP es menor al 100% dado que considero que podría haber aplicado más de una técnica multivariante para hacer un trabajo final más enriquecedor.



DIVULGACIÓN

Comunicaciones: Reunión de discusión de avances de la PP del Instituto de Administración.

Autor: Passini, Rosario Lourdes

Título: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio.

Lugar: Aula 15 de la Facultad de Ciencias Económicas

Fecha: 30/10/2019

Documentación probatoria: Cronograma de exposiciones

REUNION DE DISCUSION DE AVANCES DE LA PRACTICA PROFESIONAL DE LA

En el marco de la Práctica Profesional de la Licenciatura en Administración, se invita a docentes, estudiantes y personas interesadas a la Reunión de Discusión del Instituto de Administración que se realizará el día **Miércoles 30 de Octubre de 2019** de 20:00 a 21:30 horas en el Aula 15 de la Facultad de Ciencias Económicas. Se discutirán los avances de los siguientes trabajos:

Hora	Título del Trabajo
20:00	"Diseño e implementación de metodología para la optimización del Go To Market (G2M) basada en el Costo de Servir (C2S)" Alumno: Heluani, Matias Tutor: Medina Galvan, Marcelo Enrique
20:15	"Implementación de un nuevo sistema de información para una clínica oncológica" Alumno: Guyot, José Agustín Tutora: Rodríguez, Fernanda
20:30	"Formalización y Estandarización del Departamento de Respuestas" Alumno: Novillo, Agustín Tutor: Assaf, Diego
20:45	"Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio" Alumna: Passini, Rosario Lourdes Tutor: Medina Galván, Marcelo Enrique
21:00	"Herramientas estadísticas para la toma de decisiones" Alumno: Amone, Miguel Tutor: García, Javier

Se adjuntan los Resúmenes de los Trabajos. Esperamos contar con su presencia.

Mg. Jorge Rospide
Lic. Marcelo Medina
Instituto de Administración



Documentación probatoria: Fotografía tomada durante la exposición



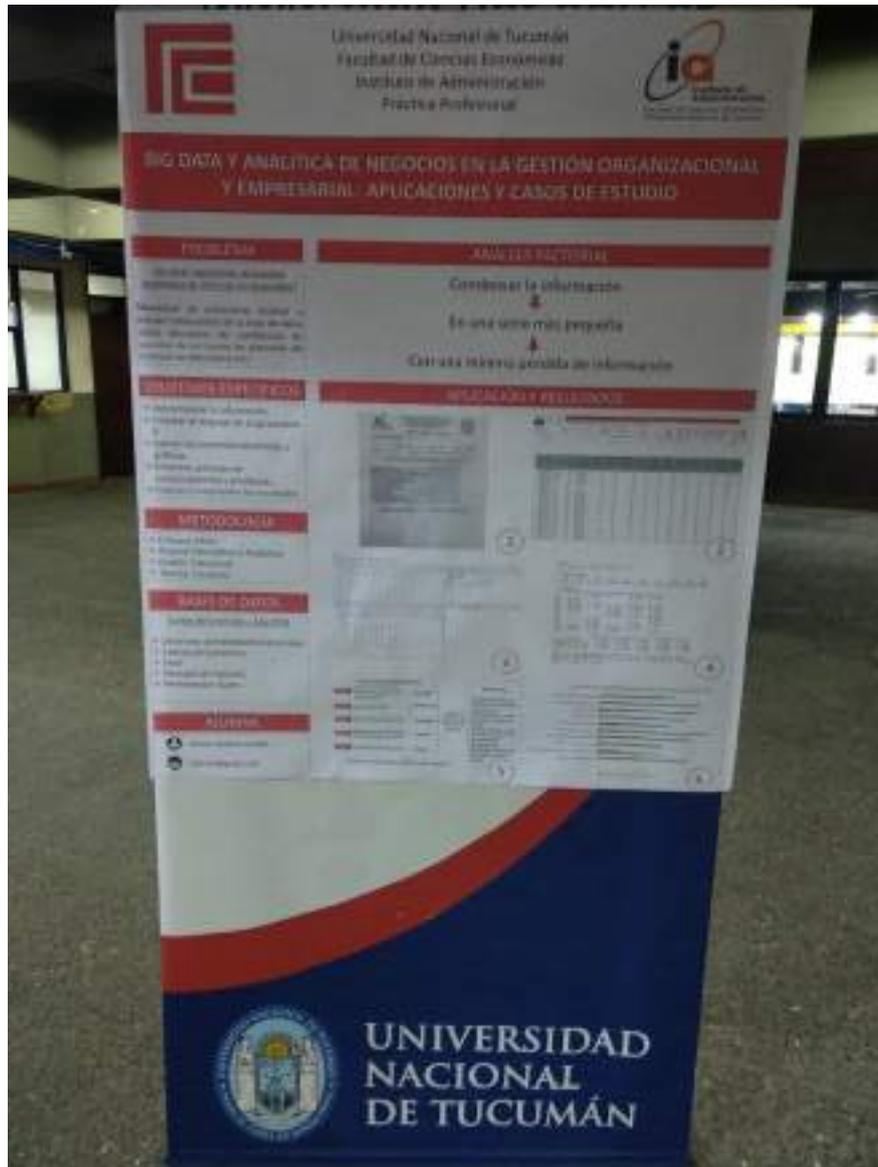
Comunicaciones: VI Muestra Académica de Trabajos de Investigación de LAD
Autor: Passini, Rosario Lourdes
Título: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial:
aplicaciones y casos de estudio.
Lugar: Aula 13 de la Facultad de Ciencias Económicas
Fecha: 14/11/2019



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



Documentación probatoria: Exhibición de poster en el hall de la Facultad de Ciencias Económicas





Documentación probatoria: Cronograma de exposiciones

inst_adm_face_unt 18 h

Práctica Profesional y Control de Gestión

Jueves

14:20	Marcopharma	Control de gestión	Aula 13
14:40	Berlin	Control de gestión	Aula 13
15:00	Andrada Hillen	PP	Aula 13
15:20	Tamargo Veronica	PP	Aula 13
15:40	Costilla Abigail	PP	Aula 13
16:00	Rentas Municipales	Control de gestión	Aula 13
16:20	Pyme farmacéutica	Control de gestión	Aula 13
16:40	Maranzana Agustina	PP	Aula 13
17:00	Passini rosario	PP	Aula 13
17:20	Hastun Josefina	PP	Aula 13
17:40	Sáncor Salud	Control de gestión	Aula 13
18:00	All frozen	Control de gestión	Aula 13
18:20	Indusnort	Control de gestión	Aula 13
18:40	JQM	Control de gestión	Aula 13
19:00	Frio banda helados	Control de gestión	Aula 13
19:20	Plasticaucho	Control de gestión	Aula 13
19:40	Corbella Luciana	PP	Aula 13
20:00	Heluani Matias	PP	Aula 13
20:20	Constanza araoz	PP	Aula 13
20:40	Centro de Salud	Control de gestión	Aula 13
21:00	Juan José Diaz	PP	Aula 13
21:20	Shell	Control de gestión	Aula 13
21:40	Alul Sofia	PP	Aula 13



CURSOS Y/O ESTADÍAS DE CAPACITACIÓN

En el mes de octubre asistí al XLVII Coloquio Argentino de Estadística donde se expusieron sobre temas relacionados a mi práctica profesional:

- Asistencia a sesiones orales de métodos multivariados donde pude observar a profesionales exponiendo la variedad de temas a los cuales se aplicaron las técnicas estadísticas.
- Asistencia a la presentación del libro “Big Data: Breve Manual para conocer la ciencia de datos que invadió nuestras vidas” donde se expuso acerca de la masiva cantidad de datos que surgen a cada instante y la importancia de utilizar técnicas, softwares y algoritmos para su procesamiento.
- Asistencia a la conferencia “Machine learning and statistical fight” para entender la importancia de la identificación de patrones complejos a partir de grandes volúmenes de datos con la finalidad de procesarlos para comprender su comportamiento, actuando de la mano con data science y trabajando a través de la estadística, métodos científicos y algoritmos.

Documentación probatoria: Certificado de asistencia



BALANCE DE LA EXPERIENCIA EN LA PRÁCTICA PROFESIONAL

Llegando el momento de finalizar mi práctica profesional puedo decir que fue una experiencia única, que se diferencia de cualquier otra materia que cursé en la facultad a lo largo de estos años. Aprendí a organizar mis propios tiempos y ser independiente en mis tareas, a hacer frente a inconvenientes que pudieron surgir y resolverlos según mi criterio de la mejor manera.

Pude entender que todos los contenidos que vemos en la facultad vuelven a surgir al momento de decidir qué temática aplicar en la práctica profesional. Además, considero que este espacio fue para mí una forma de aprendizaje que potencia y pone a prueba la habilidad con la que se identifica un problema y la destreza o desenvoltura para poder darle una solución con una amplia gama de conocimientos que nos ofrece esta carrera.

Destaco además el gran apoyo que obtuve de mi tutor, el profesor Marcelo Medina, que hasta último momento estuvo presente para contestar mis consultas y guiarme durante el transcurso de esta práctica profesional.



COMENTARIOS DEL DIRECTOR

Durante el período de la Práctica Profesional la estudiante Rosario Passini cumplió muy satisfactoriamente el tramo formativo de este espacio curricular.

Ha demostrado como competencias destacables el pensamiento analítico, la proactividad y la autodisciplina.

Se destaca la amplia lectura de bibliografía y la participación en el Coloquio de la Sociedad Argentina de Estadística celebrado en la FACE.

Ha participado en las reuniones de tutorías, en reuniones de discusión de avances de la Práctica Profesional y la Muestra Académica organizada en por el Instituto de Administración.

Como contribución a la Práctica Profesional, el trabajo elaborado por Rosario representa un aporte a la aplicación de la Analítica de Negocios para la explotación de datos de encuestas de satisfacción de usuarios y a la generación de conocimiento para colaborar en pos de una mejor gestión.

El desempeño de Rosario se considera altamente satisfactorio.

Firma del Director_____

Fecha

Aclaración_____



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



DOCUMENTACIÓN PROBATORIA



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



PLAN DE TRABAJO



*Plan de Trabajo: “Big Data
y Analítica de Negocios en
la Gestión Organizacional
y Empresarial: aplicaciones
y casos de estudio”*

Práctica Profesional Licenciatura en Administración





BIG DATA Y ANALÍTICA DE NEGOCIOS EN LA GESTIÓN ORGANIZACIONAL Y EMPRESARIAL: APLICACIONES Y CASOS DE ESTUDIO

Passini, Rosario Lourdes

rlpassini@gmail.com

RESUMEN

Poder encontrar el beneficio real de la utilización de los datos es una necesidad fundamental de todas las empresas y organizaciones en la actualidad, los datos solamente serán valiosos si se pueden procesar, comprender, analizar y utilizar. Hoy en día se cuenta con una gran cantidad de información disponible que va creciendo y almacenándose cada vez más de manera exponencial, la gestión y análisis de estos enormes volúmenes de datos es lo que se denomina como Big Data, a la vez que se utiliza la Analítica de negocios (Business Analytics, BA) como práctica de exploración iterativa y metódica de los datos existentes para realizar un análisis estadístico de la información que ayude a obtener respuestas a miles de preguntas esenciales para el correcto funcionamiento y para la toma de decisiones en los negocios, colaborando además con la identificación y aprovechamiento de nuevas oportunidades.

El presente proyecto prevé la utilización de herramientas cualitativas y técnicas estadísticas para gestionar la información disponible en las organizaciones haciendo uso de la minería de datos para encontrar patrones ocultos, repetitivos o tendencias que expliquen el comportamiento de los datos con la finalidad de ofrecer resultados que sean útiles en un determinado contexto.

Palabras Claves: Analítica de Negocios, Big Data, Minería de Datos, Estadística



INDICE

1- MODALIDAD	15
2- INTRODUCCIÓN	15
3- PROBLEMA.....	15
4- OBJETIVOS.....	16
5- MARCO TEÓRICO	16
¿Qué es el análisis univariado, bivariado y multivariado?	16
Taxonomía de las herramientas estadísticas	16
¿Qué significa Big Data?	18
¿Qué es la minería de datos?	18
¿En qué consiste la analítica de negocios?	19
6- MARCO METODOLÓGICO.....	19
7- CARGA HORARIA	19
8- CRONOGRAMA DE ACTIVIDADES.....	20
9- BIBLIOGRAFÍA	21



1- MODALIDAD

Teniendo en cuenta las cuatro posibilidades ofrecidas por la Facultad de Ciencias Económicas para el desarrollo de la Práctica Profesional (PP), el presente trabajo será realizado bajo la modalidad de “Participación en Proyectos de investigación”, se opta por el proyecto “Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio” donde se llevarán a cabo actividades de la investigación propiamente dicha como ser: lectura de material relacionado, recolección de datos, aplicación y análisis de herramientas estadísticas, evaluación de los resultados.

2- INTRODUCCIÓN

Analizar datos permite a las organizaciones enriquecer la información que tienen disponible con la finalidad de poder comprender el porqué de determinados comportamientos y mejorar la capacidad para poder predecirlos, el análisis y la gestión de esta información facilitan la toma de decisiones y la creación de ventajas competitivas. Por esta razón, resulta necesario contar con herramientas que ayuden a examinar los datos existentes que posibiliten establecer ciertas relaciones y comparaciones entre diversas variables que permitan al interesado adelantarse al futuro.

En este trabajo se hace uso de una base de datos que contiene una recopilación de encuestas de satisfacción de usuarios de los cursos de extensión del Instituto de Administración en la Facultad de Ciencias Económicas de la UNT donde se pretende aplicar herramientas tanto cualitativas como cuantitativas para la gestión de la información. Con respecto a los métodos cuantitativos se prevé el empleo del análisis univariado, bivariado y/o multivariado para la recopilación y evaluación de los datos. Como método cualitativo se dispone la utilización de la etnografía, estudio de casos e investigación documental. El cumplimiento de este proyecto estima el paso por las siguientes etapas: lectura del material bibliográfico, revisión de la base de datos y definición del problema, recolección de la información, identificación y aplicación de técnicas, análisis e interpretación de los resultados, presentación de los resultados.

3- PROBLEMA

Problema general:

¿Se están explotando las grandes cantidades de información disponible?

Problema específico:



Necesidad de estructurar, analizar y extraer conclusiones de la base de datos sobre encuestas de satisfacción de usuarios de los cursos de extensión del Instituto de Administración.

4- OBJETIVOS

Los objetivos que se persiguen en este trabajo final son los siguientes:

Objetivo general: Explotar los datos contenidos en las encuestas de satisfacción de los cursos de extensión para extraer información relevante que se pueda procesar y analizar hasta obtener conclusiones concretas.

Objetivos específicos:

- ❖ Descomponer la información contenida en la base de datos para identificar las variables con las cuales se va a trabajar y el comportamiento que se quiere predecir.
- ❖ Emplear el lenguaje de programación R para la minería de datos.
- ❖ Aplicar herramientas estadísticas y gráficas para clasificar, agrupar y asociar la información.
- ❖ Encontrar patrones de comportamiento y tendencias.
- ❖ Hacer uso de la información obtenida para evaluar y comprender los resultados.

5- MARCO TEÓRICO

¿Qué es el análisis univariado, bivariado y multivariado?

Se trata del análisis cuantitativo de datos según el número de variables que se quiere estudiar. Cuando se analiza una única variable y se emplea la estadística descriptiva para calcular medidas de tendencia central como un promedio, una mediana y una moda; así también cuando se estima la dispersión o variabilidad como la varianza o la desviación estándar, se está empleando un análisis univariado. Ahora bien, cuando se examinan dos variables con la finalidad de conocer la causalidad se puede hablar del análisis bivariado. Por último, estamos en presencia del análisis multivariado cuando se estudian más de dos variables a la vez.

Taxonomía de las herramientas estadísticas

Se presenta una clasificación de los métodos multivariantes en la siguiente figura:

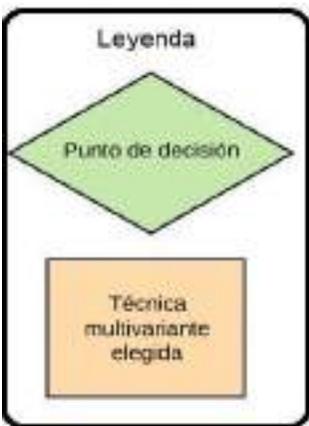
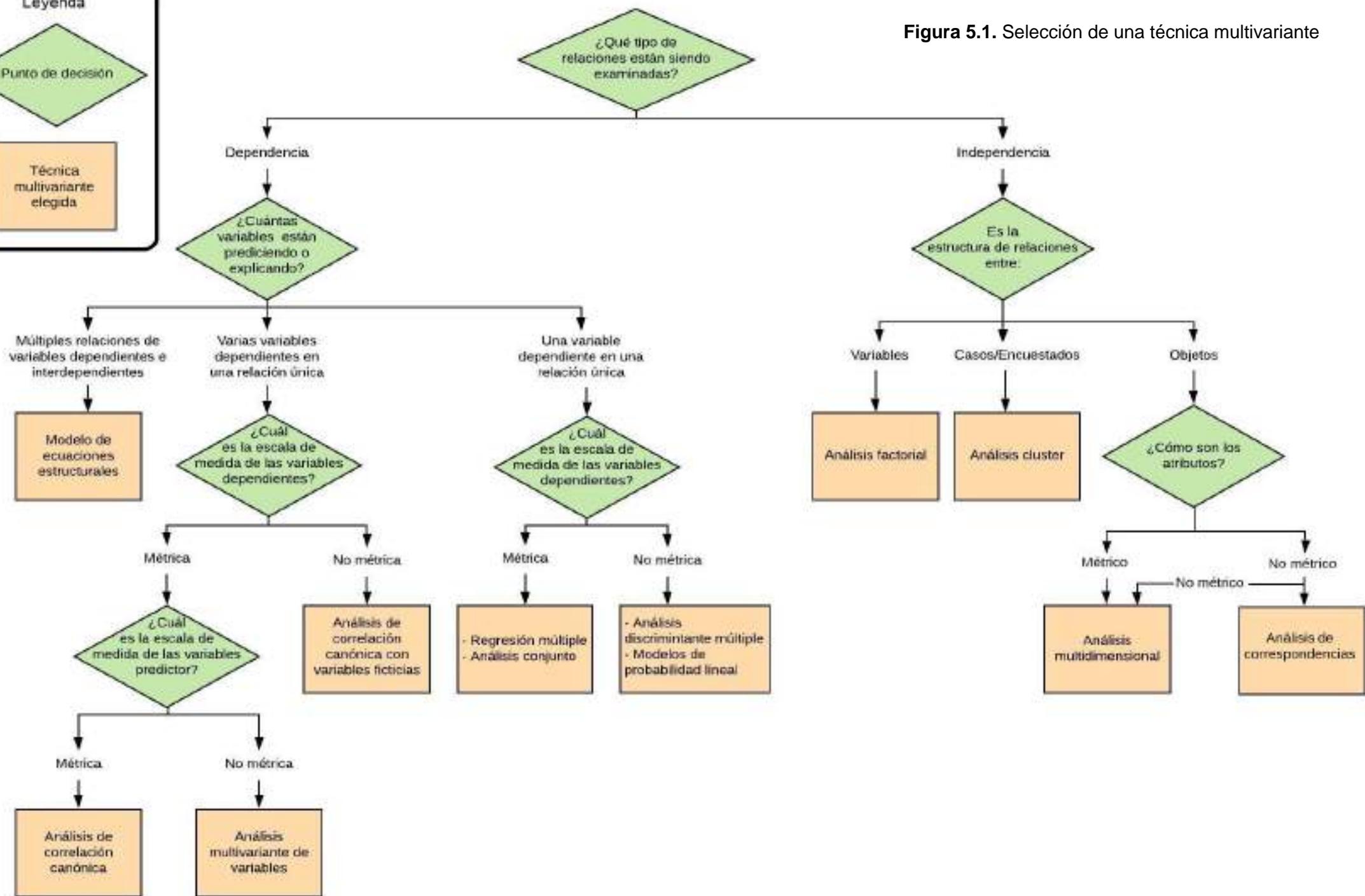


Figura 5.1. Selección de una técnica multivariante





Esta clasificación se basa en tres juicios que el analista debe hacer sobre el objeto a investigar y la naturaleza de los datos: (1) ¿pueden dividirse las variables en dependientes o independientes basándose la clasificación en alguna teoría? (2) Si puede hacerse, ¿cuántas de estas variables son tratadas como dependientes en un análisis simple? (3) ¿Cómo son las variables medidas? La selección de la técnica multivariante apropiada depende de las respuestas a estas tres cuestiones.

¿Qué significa Big Data?

Se denomina Big Data a la gestión y análisis de enormes volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software habitualmente utilizadas para la captura, gestión y procesamiento de datos. Las ventajas del Big Data son las siguientes:

- ❖ Permite obtener respuestas más completas, ya que dispone de mayor cantidad de información.
- ❖ La disponibilidad de respuestas más completas significa una mayor fiabilidad de datos, lo que implica un enfoque completamente distinto a la hora de abordar problemas.

Las cinco características más importantes del Big Data (5 Vs) son:

- ❖ Volumen: El concepto de volumen es muy variable y cada día que pasa se eleva lo que podemos considerar como grandes volúmenes de datos.
- ❖ Variedad: Infinidad de tipos de datos se agrupan dispuestos a ser tratados y es por ello que frente a esa variedad aumenta el grado de complejidad tanto en el almacenamiento como en su análisis.
- ❖ Velocidad: Se refiere a la rapidez con que los datos se reciben, se procesan y se toman decisiones a partir de ellos.
- ❖ Veracidad: Se refiere a la confianza de los datos o información obtenida.
- ❖ Valor: El valor es relativo, es decir, la importancia del dato se la dará cada negocio, sus objetivos y las estrategias que aplique para la medición.

¿Qué es la minería de datos?

La minería de datos es el proceso de búsqueda en grandes bases de datos para encontrar información útil que sirva para la toma de decisiones. Se puede entender como la tecnología y software utilizado de manera automática o semiautomática para encontrar patrones de comportamiento dentro de la base de datos. Básicamente, el datamining surge para intentar ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y, en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

Utiliza el análisis matemático para deducir cada tendencia, y su importancia yace en que estos patrones no se pueden detectar mediante la exploración tradicional



de datos, ya que son muy complejas relaciones o porque la cantidad de datos es demasiada.

¿En qué consiste la analítica de negocios?

El análisis de negocios es un enfoque basado en datos que permite a las personas tomar las mejores decisiones, implica la utilización de grandes cantidades de datos, lo cual significa que la tecnología de la información relacionada con la gestión de datos es muy importante. Se utiliza el análisis estadístico y cuantitativo para analizar los datos y proporcionar información útil a quienes toman decisiones. El análisis de negocios suele dividirse en tres categorías: descriptivo, predictivo y prescriptivo.

El análisis descriptivo implica el estudio y la consolidación de los datos históricos de un negocio, ayuda a medir como se ha desempeñado una organización en el pasado y cómo lo está haciendo en el momento actual. El análisis predictivo está dirigido a pronosticar los resultados futuros, con base en los patrones de los datos históricos (para este propósito se utilizan ampliamente los modelos estadísticos y matemáticos). El análisis prescriptivo implica el uso de métodos de optimización para ofrecer nuevas y mejores formas de operar, con base en los objetivos específicos del negocio.

6- MARCO METODOLÓGICO

Se define así al grupo de mecanismos o procedimientos racionales, empleados para el logro de los objetivos que dirige una investigación.

La población de estudio son las personas que asistieron a los cursos de extensión del instituto de administración dictados en la Facultad de Ciencias Económicas de la UNT en el año 2018 y parte del 2019.

El enfoque llevado a cabo en este trabajo es del tipo mixto, es cuantitativo por el hecho de que se trabaja con datos cuantificables, haciendo uso de herramientas informáticas, estadísticas y matemáticas para el análisis de los mismos, y es cualitativo por la utilización de la etnografía, estudio de casos e investigación documental.

El alcance es descriptivo y predictivo, dado que se busca definir variables, cuantificarlas y ver como se relacionan entre ellas.

El diseño de la investigación sigue un estudio transversal porque se analiza la relación de las variables en un solo punto en el tiempo.

La técnica utilizada para la recolección de datos es la encuesta a los participantes de cursos de extensión del instituto de investigación.

7- CARGA HORARIA

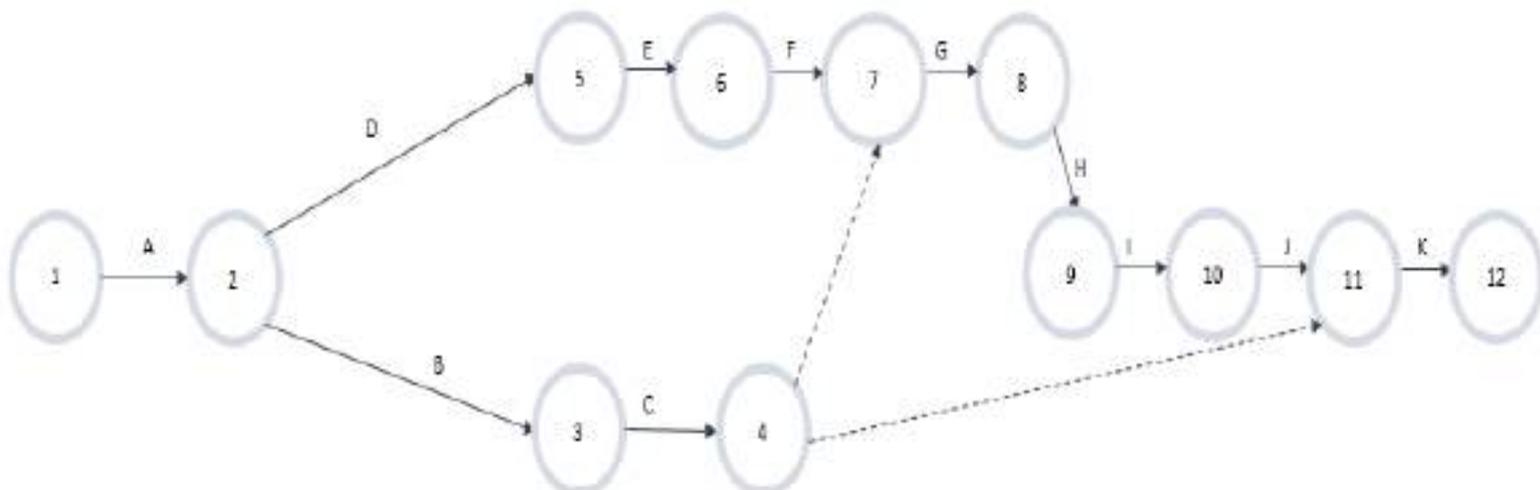
Para este proyecto se prevé una carga horaria de 400 horas hasta su finalización. La utilización de estas horas se espera emplearlas de la siguiente manera:

- ❖ Asimilación de conceptos teóricos: 160 horas.
- ❖ Aplicación de herramientas prácticas: 200 horas.
- ❖ Elaboración del informe final: 40 horas.



8- CRONOGRAMA DE ACTIVIDADES

ETAPAS	CÓDIGO	ACTIVIDADES	PRECEDENCIA	DURACIÓN (HS)
Asimilación de conceptos teóricos	A	Lectura del material bibliográfico	-	80
	B	Recopilación de la información	A	40
	C	Construcción del marco teórico	B	40
Aplicación de herramientas prácticas	D	Análisis de base de datos	A	40
	E	Definición del problema	D	20
	F	Recolección de datos	E	40
	G	Aplicación de técnicas estadísticas en R	C,F	60
	H	Identificación de patrones de comportamiento	G	40
Elaboración del informe final	I	Interpretación de los resultados	H	20
	J	Planteo de conclusiones y recomendaciones	I	10
	K	Elaboración de informe final	C,J	10





9- BIBLIOGRAFÍA

- J. F. Hair, Jr., R. E. Anderson, R. L. Tatham, W. C. Black (1999). Análisis Multivariante, 5° edición. Iberia, Madrid. Prentice Hall.
- Brian S. Everitt and Torsten Hothorn (2005). A Handbook of Statistical Analyses Using R. Recuperado de <https://static1.squarespace.com/static/57af6d69b3db2b8d0eade247/t/580a6f849f74564489f129dc/1477078927168/R2.pdf>
- Barry Render, Ralph Stair, Michael Hanna, Trevor Hale (2016). Métodos cuantitativos para los negocios, 12° edición. México. Pearson.
- Daniel Peña (2002). Análisis multivariante de datos. Mc Graw Hill Interamericana



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



INFORMES DE AVANCE



Primer Informe de Avance – Práctica Profesional

Título: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio

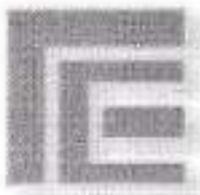
Alumna: Passini, Rosario Lourdes

Tutor: Medina Galván, Marcelo Enrique

Periodo: Agosto - Septiembre

Actividades realizadas:

- Lectura de los libros "Análisis multivariante" y "Análisis multivariante de datos" con la finalidad de poder armar el marco teórico e identificar la técnica multivariante a utilizar en el presente trabajo.
- Lectura de un capítulo seleccionado del libro "Métodos cuantitativos para los negocios" para entender el concepto de la analítica de negocios, sus categorías e implicancia.
- Lectura de artículos en internet sobre big data, minería de datos, analítica de negocios, técnicas multivariantes.
- Seguimiento de tutoriales y lectura de instructivos sobre el software RStudio para entender su funcionamiento.
- Recolección de las bases de datos de las encuestas de los cursos de extensión del año 2019 con las cuales se procederá a trabajar.



Segundo Informe de Avance – Práctica Profesional

Título: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio

Alumna: Passini, Rosario Lourdes

Tutor: Medina Galván, Marcelo Enrique

Periodo: Octubre

Actividades realizadas:

- Luego de haber leído la bibliografía acerca de los métodos multivariantes y habiendo optado por la utilización del análisis factorial para ser aplicado en el presente trabajo, se continuó con buscar información adicional sobre esta técnica estadística para poder comprender aún más su aprovechamiento.
- Asistencia a XLVII Coloquio Argentino de Estadística
- Descomposición de la información contenida en las bases de datos para filtrar las variables con las cuales se iba a trabajar y dejar de lado las que no son necesarias para el análisis.
- Se trabajó con el software RStudio cargando la base de datos del curso de extensión "Desarrollo de habilidades gerenciales" con la finalidad de aplicar un análisis factorial a los datos y confirmar si esta técnica estadística es la apropiada.
- Obtención de los resultados parciales e interpretación de los mismos.
- Elaboración del resumen y diapositivas para ser presentados en la reunión de discusión de avances de la PP.
- Exposición en la última reunión de discusión del día 30/10



Tercer Informe de Avance – Práctica Profesional

Título: Big Data y Analítica de Negocios en la Gestión Organizacional y Empresarial: aplicaciones y casos de estudio

Alumna: Passini, Rosario Lourdes

Tutor: Medina Galván, Marcelo Enrique

Período: Noviembre

Actividades realizadas:

- Reunión con mi tutor el día 31/10 para la revisión de los resultados parciales, consulta de dudas y decidir cómo seguir trabajando con las bases de datos restantes.
- Lectura de artículos acerca de la diferencia entre análisis factorial exploratorio y análisis factorial confirmatorio.
- Revisión de los videos tutoriales de R para aprender a graficar los datos.
- Aplicación del análisis factorial en las 4 bases de datos que faltaban estudiar a través de RStudio.
- Elaboración del poster, documento y edición de diapositivas para presentar en la VI Muestra Académica de Trabajos de Investigación de la Licenciatura en Administración.
- Confección parcial del documento final de la Práctica Profesional.
- Reunión con mi tutor el día 07/11 para controlar los resultados obtenidos.



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



TRABAJO FINAL



Resultados del Trabajo de Campo

Práctica Profesional Licenciatura en Administración





BIG DATA Y ANALÍTICA DE NEGOCIOS EN LA GESTIÓN ORGANIZACIONAL Y EMPRESARIAL: APLICACIONES Y CASOS DE ESTUDIO

Passini, Rosario Lourdes

rlpassini@gmail.com

RESUMEN

Hoy en día no resulta suficiente contar con una base de datos disponible para reforzar la toma de decisiones, se requiere también de su análisis a fondo. Esta temática denominada analítica de negocios toma datos, los filtra, analiza y convierte en información que puede usarse para tomar decisiones, encontrar conclusiones y en ciertos casos hacer predicciones a partir de un conjunto de datos. Los beneficios de este análisis son incontables.

Con las tecnologías apropiadas podemos capturar, almacenar, procesar y hasta presentar los datos de una manera que se pueda “consumir” más fácilmente.

La analítica de datos permite a las personas conocer lo que pasa con sus negocios, conocer las falencias y fortalezas que pueden surgir para poder transformarlas y encaminar sus intereses hacia el éxito.

Con la finalidad de poder estudiar los fenómenos que abarcan la medición de varias variables se puede utilizar un conjunto de análisis estadísticos llamados “Métodos Multivariantes”, este trabajo hace uso de la técnica multivariante denominada Análisis factorial aplicada sobre una serie de datos obtenidos mediante encuestas de satisfacción de 5 cursos de extensión dictados durante el año 2019 en la Facultad de Ciencias Económicas.

Palabras Claves: Métodos multivariantes, Analítica de negocios, Análisis factorial



INDICE

1- INTRODUCCIÓN	30
2- METODOLOGÍA	30
3- MARCO TEÓRICO	31
Big Data	31
Data Science.....	32
Análisis Multivariado.....	34
Análisis Factorial	35
4- APLICACIÓN DEL ANÁLISIS FACTORIAL EN RStudio	36
Paso a paso de la aplicación de la técnica	36
5- INTERPRETACIÓN DE LOS RESULTADOS.....	42
5.1 Curso: Desarrollo de Habilidades Gerenciales	42
5.2 Curso: Excel	44
5.3 Curso: Mercado de Capitales	46
5.4 Curso: Cadena de Suministros	48
5.5 Curso: Metodologías Ágiles	50
6- CONCLUSIONES	52



1- INTRODUCCIÓN

La encuesta de satisfacción es una excelente herramienta para poder visualizar mejoras que se necesitan hacer en una organización y para escuchar las necesidades y demandas de los usuarios o consumidores.

El Instituto de Administración de la Facultad de Ciencias Económicas posee los datos obtenidos de cada uno de los cursos de extensión que se dictan a lo largo del año y resulta necesario su análisis para conocer el desempeño de la oferta de estos talleres y el grado de satisfacción de los usuarios para poder ofrecer oportunidades de mejora en el futuro.

Para poder llevar a cabo este trabajo se procede a utilizar las encuestas de satisfacción que posee el Instituto de Administración sobre los cursos de extensión dictados en la Facultad de Ciencias Económicas de la UNT.

El total de cursos de extensión que se analizan a lo largo de esta práctica profesional son 5, ellos son: Desarrollo de habilidades gerenciales, Excel, Mercado de capitales, Cadena de suministros y Metodologías ágiles. Estos cursos se dictaron durante el presente año y los datos de cada uno se encuentran contenidos en planillas de Excel.

El propósito del análisis de estos datos es tratar de descubrir qué piensan los alumnos, cuál es su opinión sobre los cursos y qué aspectos del mismo perciben como valiosos, de manera que se pueda tener un claro conocimiento acerca de la propuesta de valor ofrecida para cada uno de los cursos.

2- METODOLOGÍA

Se define así al grupo de mecanismos o procedimientos racionales, empleados para el logro de los objetivos que dirige una investigación.

La población de estudio son las personas que asistieron a los 5 cursos de extensión del instituto de administración (Desarrollo de habilidades gerenciales, Excel, Mercado de capitales, Cadena de suministros, Metodologías ágiles), dictados en la Facultad de Ciencias Económicas de la UNT en el transcurso del año 2019.

El enfoque llevado a cabo en este trabajo es del tipo mixto, es cuantitativo por el hecho de que se trabaja con datos cuantificables, haciendo uso de herramientas informáticas, estadísticas y matemáticas para el análisis de los mismos, y es cualitativo por la utilización de la etnografía, estudio de casos e investigación documental.

El alcance es descriptivo y predictivo, dado que se busca definir variables, cuantificarlas y ver como se relacionan entre ellas. Se busca especificar



propiedades, características y rasgos importantes del fenómeno analizado, además se intentan predecir tendencias en los comportamientos.

El diseño de la investigación sigue un estudio transversal porque se analiza la relación de las variables en un solo punto en el tiempo. Este corte transversal se lleva a cabo en una sola instancia donde las variables se recolectan en un determinado punto en el tiempo y no cambian durante el periodo de investigación.

La técnica utilizada para la recolección de datos es la encuesta de satisfacción de usuarios compuesta de un conjunto de preguntas dirigidas a la población bajo estudio que son los participantes de los 5 cursos de extensión nombrados anteriormente, la finalidad de esta técnica es obtener los datos necesarios para su posterior análisis.

3- MARCO TEÓRICO

Big Data

¿En que consiste?

Actualmente, se generan muchos datos y las empresas exigen saber no sólo lo que sucede en la actualidad, sino también lo que va a pasar en un futuro, requiriendo niveles de servicio mucho más exigentes que hace unos años. Es precisamente en esta tesitura donde se encuentra Big Data hoy.

¿Por qué es tan importante?

La importancia de Big Data radica que en que éste impacta tanto en la industria, como en el negocio e incluso en nuestra sociedad y además ofrece una ventaja competitiva considerable.

En efecto, es precisamente en ese tipo de datos donde las empresas han detectado que se encierra mayor valor. Hoy en día, para muchas empresas puede llegar a ser más importante detectar al cliente que más influye al resto de posibles compradores, que al que mayor volumen de compra realiza.

En la actualidad, la cantidad de datos que se generan es abismal y de una casuística extremadamente compleja para su análisis. Como hemos comentado, las empresas cada vez exigen que el análisis sea lo más cercano posible al tiempo real. Y en Big Data está la clave, al traducirse el mismo en las variables de velocidad, variedad y volumen que requiere el mercado actualmente.



Figura 3.1: Las 7 V's del big data

Big Data requiere Big Analytics

Estamos acostumbrados a saber lo que pasó, pero hoy nos interesa más conocer lo que pasará: si la empresa, con su producto o servicio, seguirá siendo el gusto de los consumidores o si se hablará de ella bien o mal en las redes sociales; aspectos todos ellos que requieren de nuevos modelos de análisis mucho más complejos que lo que se podía experimentar hasta ahora.

Analytics permite la aplicación de procesos matemáticos complejos. Sus técnicas hacen posible trabajar con los datos al nivel de granularidad más bajo disponible, en crudo, y empleando para ello modelos mucho más ágiles que los actuales modelos de BI.

Data Science

Es la ciencia que estudia la extracción de conocimiento a partir de los datos. Data science es el procedimiento de obtener información valiosa de los datos y está apareciendo debido a la necesidad de trabajar con conjuntos inmensos de datos (conocidos como Big data), formados tanto por datos estructurados como por datos semiestructurados o desestructurados. Estos conjuntos proceden de los datos generados por los dispositivos electrónicos, las redes sociales y la web.

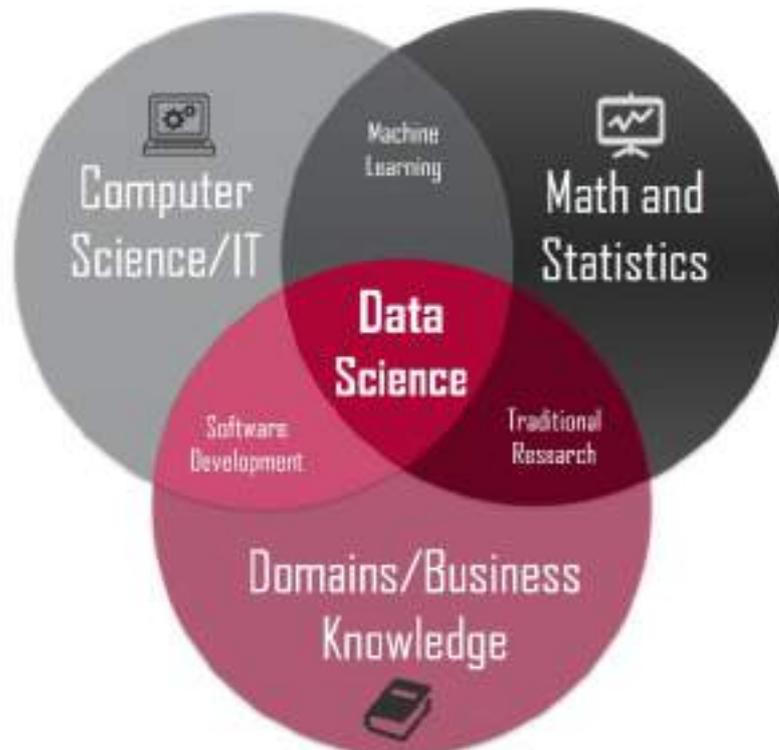


Figura 3.2: Conceptos relacionados al Data Science

Tipos de datos

- Datos estructurados: Los datos estructurados tienen perfectamente definido la longitud, el formato y el tamaño de sus datos.
- Datos no estructurados: Los datos no estructurados se caracterizan por no tener un formato específico. Se almacenan en múltiples formatos como documentos PDF o Word, correos electrónicos, ficheros, etc.
- Datos semiestructurados: son una mezcla de los dos anteriores no presenta una estructura perfectamente definida como los datos estructurados pero si presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención, como por ejemplo los formatos HTML, XML o JSON.



Data Science y Big Data

El término Big Data se suele relacionar con Ciencia de Datos, pues esta suele ser su fuente de información para análisis; La ciencia de datos logra analizar los grandes conjuntos de datos desordenados e incompletos, para llegar a hallazgos que impulsan decisiones sobre operaciones y productos.

El proceso que sigue un científico de datos para responder cuestiones que se le plantean se puede resumir en estos pasos:

- Extraer datos, independientemente de la fuente y de su volumen.
- Limpiar los datos, para eliminar lo que pueda sesgar los resultados.
- Procesar los datos usando métodos estadísticos como inferencia estadística, modelos de regresión, pruebas de hipótesis, etc.
- Diseñar experimentos adicionales en caso de ser necesario.
- Crear visualizaciones graficas de los datos relevantes de la investigación



Figura 3.3: Big Data y su análisis

Análisis Multivariado

Es el conjunto de métodos estadísticos, cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes, en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado.



Objetivos del análisis

Pueden sintetizarse en dos:

- 1) Proporcionar métodos, cuya finalidad es el estudio conjunto de datos multivariantes, que el análisis estadístico uni y bidimensional es incapaz de conseguir.
- 2) Ayudar a analizar e investigar, a tomar decisiones óptimas, en el contexto en el que se encuentre, teniendo en cuenta la información disponible por el conjunto de datos analizado.

Algunos de los métodos multivariantes más usados son los siguientes:

- Análisis factorial
- Análisis de regresión múltiple
- Análisis discriminante
- Análisis multivariante de la varianza
- Análisis conjunto
- Análisis de correlación canónica
- Análisis cluster
- Análisis multidimensional
- Análisis de correspondencias

Análisis Factorial

El Análisis Factorial Exploratorio (AFE) [Exploratory Factor Analysis; EFA] es una técnica estadística que permite explorar con mayor precisión las dimensiones subyacentes, constructos o variables latentes de las variables observadas, es decir, las que observa y mide el investigador.

El análisis factorial consiste en una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numerosos de variables. Esos grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

Cuando recogemos un gran número de variables de forma simultánea, como por ejemplo en un cuestionario de satisfacción laboral, podemos estar interesados en averiguar si las preguntas del cuestionario se agrupan de alguna forma característica. Aplicando un análisis factorial a las respuestas de los sujetos podemos encontrar grupos de variables con significado común y conseguir de esta manera reducir el número de dimensiones necesarias para explicar las respuestas de los sujetos.

El análisis factorial es, por tanto, una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos.



A diferencia de lo que ocurre en otras técnicas como el análisis de varianza o el de regresión, en el análisis factorial todas las variables del análisis cumplen el mismo

papel: todas ellas son independientes en el sentido de que no existe a priori una dependencia conceptual de unas variables sobre otras.

4- APLICACIÓN DEL ANÁLISIS FACTORIAL EN RStudio

Luego de haber analizado las cinco bases de datos de los cursos de extensión dictados en el presente año (Mercado de capitales, Desarrollo de habilidades gerenciales, Cadena de suministro, Metodologías ágiles y Excel) y las variables que contienen las mismas, se procede a optar por la utilización del análisis factorial exploratorio para su estudio.

El propósito del análisis factorial es encontrar una manera de condensar la información contenida en una serie de variables originales en una serie más pequeña de dimensiones compuestas o valores teóricos (factores) nuevos con una mínima pérdida de información. El análisis factorial resume los datos en unas pocas dimensiones al condensar un gran número de variables en un conjunto más pequeño de factores.

Se procede a mostrar un ejemplo del paso a paso de la técnica empleada utilizando como ejemplo el curso “Desarrollo de Habilidades Gerenciales”. En la siguiente sección (5- INTERPRETACIÓN DE LOS RESULTADOS) se puede ver la interpretación de las salidas obtenidas para cada uno de los 5 cursos de extensión

Paso a paso de la aplicación de la técnica

Primero: Se muestra a continuación la base de datos del curso “Desarrollo de habilidades gerenciales” (DHG) en su formato Excel, esta base de datos se obtiene al volcar la información que está incluida en cada una de las encuestas de satisfacción para este curso:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Edad	Sexo	Carrera	Trabajo	Estado Civil	Cantidad de materias aprobadas	Utilidad percibida del curso	Diferencia de conocimiento entre inicio y final	Temas tratados en el curso	Cantidad y profundidad del docente	Atino y claridad del taller	Cantidad de horas del taller	Horario del taller	Lugar del taller	Motivación del participante	Calificación del taller en general
1	21	Masculino	IAE	NO	Estudiante	14	3	4	4	4	5	5	4	3	4	4
2	26	Femenino	BBBA	SI	Graduado		3	4	5	5	5	5	5	5	5	5
3	26	Masculino	IAE	NO	Estudiante	10	3	3	5	5	5	5	5	5	5	5
4	25	Femenino	IAE	SI	Estudiante	27	4	5	4	5	5	5	5	4	5	5
5	18	Masculino	SI				4	4	4	4	4	4	4	4	4	4
6	48	Femenino	CPV	SI	Graduado		4	3	4	4	5	4	5	5	4	4
7	22	Femenino	IAE	NO	Estudiante	10	3	3	3	3	3	3	3	3	3	3
8	27	Femenino	CPV	SI	Graduado		4	4	3	5	5	5	4	5	5	5
9	20	Masculino	CPV	NO	Estudiante	10	3	4	5	5	5	4	4	5	5	5
10	18	Femenino	IAE	NO	Estudiante	10	3	4	3	3	3	3	3	3	3	3
11	32	Femenino	CPV	SI	Graduado		3	4	4	5	5	4	4	4	5	5
12	18	Masculino	CPV	SI	Estudiante	16	3	4	3	5	5	5	5	5	5	5
13	28	Femenino	IAE	SI	Graduado		3	4	3	3	4	3	3	3	4	4
14	19	Femenino	IAE	SI	Graduado		3	4	4	5	5	4	5	5	5	4
15	40	Femenino	CPV	SI	Graduado		4	4	4	5	5	5	5	5	5	5
16	29	Masculino	CPV	SI	Graduado		3	4	4	4	4	5	5	5	5	4
17	30	Femenino	IAE	SI	Estudiante	20	4	4	4	4	4	4	4	4	4	4
18	48	Femenino	CPV	SI	Graduado		3	3	3	4	5	4	5	5	4	4
19	32	Femenino	IAE	SI	Graduado		4	4	4	5	5	5	5	5	4	5

Figura 4.1 Descripción de las variables de la base de datos del curso DHG

Segundo: Para poder realizar el análisis en el software RStudio se deben filtrar las variables a utilizar y renombrarlas para que sea visiblemente más cómodo al momento de trabajar con ellas. En este caso se hace uso de las últimas 10 variables que representan los atributos del curso que los alumnos debieron calificar en una escala del 1 al 5 (donde 1 = malo y 5 = excelente)



	A	B	C	D	E	F	G	H	I	J
1	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
2	5	4	4	4	5	5	4	3	4	4
3	5	4	5	5	5	5	5	5	5	5
4	5	3	5	5	5	5	5	5	5	5
5	4	5	4	5	5	3	5	4	5	5
6	4	4	4	4	4	4	4	4	5	4
7	4	3	4	4	5	4	5	3	4	4
8	5	5	5	5	5	5	5	5	5	5
9	4	4	5	5	5	3	4	5	5	5
10	5	4	5	5	5	4	4	5	5	5
11	5	4	5	5	5	5	5	5	5	5
12	5	4	4	5	5	4	4	4	5	5
13	5	4	5	5	5	5	5	5	5	5
14	5	3	4	5	5	4	5	5	4	4
15	5	4	4	5	5	4	5	5	5	4
16	4	4	4	5	5	5	5	5	5	5
17	5	4	4	4	4	3	5	5	5	4
18	4	4	4	4	4	4	4	4	4	4
19	5	3	5	4	3	4	5	5	4	4
20	4	4	4	5	5	5	5	5	4	5
21	5	3	4	5	5	5	4	5	5	4
22	5	5	5	5	5	5	4	5	5	5
23	4	3	3	3	4	3	4	4	4	4
24	4	3	4	4	5	5	5	5	5	4
25	5	3	5	5	5	5	5	5	5	5

Tabla 4.1. Matriz con datos filtrados necesarios para emplear el análisis factorial

Tercero: Obtenida la matriz de 27x10 (27 observaciones y 10 variables) se continúa con su carga en el software para la aplicación de la técnica estadística elegida.

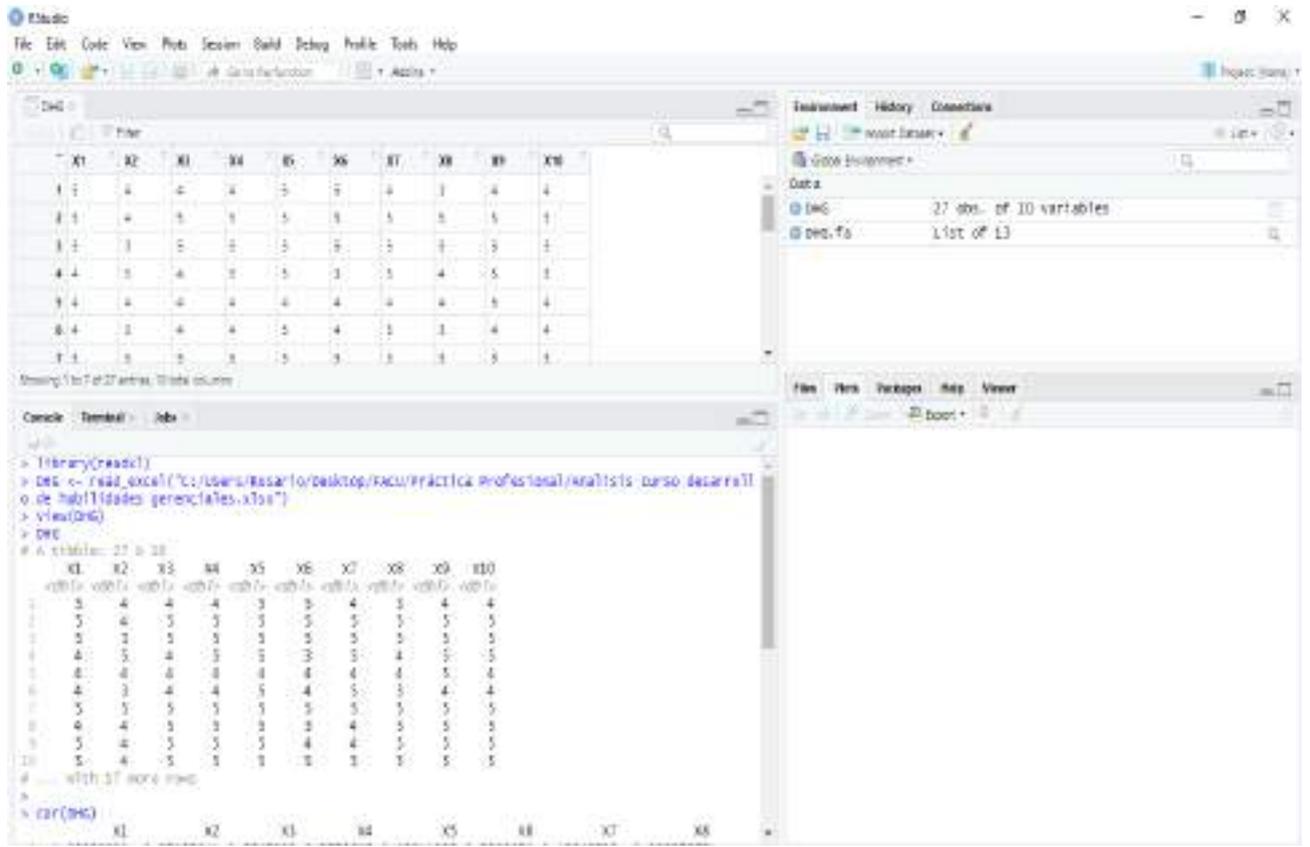


Figura 4.2. Captura del Software RStudio con la matriz cargada

Cuarto: Empleando el análisis factorial con 5 factores utilizando el comando “factanal” se obtiene el siguiente resultado:



```
call:
factanal(x = DHG, factors = 5)

Uniquenesses:
  x1  x2  x3  x4  x5  x6  x7  x8  x9  x10
0.363 0.595 0.290 0.251 0.005 0.404 0.768 0.005 0.491 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
x1      0.163          0.769  0.102
x2  0.623
x3  0.419  0.303          0.614  0.253
x4  0.323  0.515  0.529  0.266  0.169
x5  0.213          0.956          0.174
x6          0.308  0.354  0.609
x7          0.163          0.449
x8          0.932          0.207  0.287
x9  0.337  0.466  0.329  0.261
x10 0.871  0.242  0.254          0.331

      Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings  1.593  1.559  1.482  1.289  0.901
Proportion Var  0.159  0.156  0.148  0.129  0.090
Cumulative Var  0.159  0.315  0.463  0.592  0.682

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 2.11 on 5 degrees of freedom.
The p-value is 0.834
>
```

Figura 4.3. Salida del software para el curso DHG

La razón por la cual se elige una solución con 5 factores se debe a que dicha solución contiene el mayor porcentaje de varianza acumulada explicada de los datos.

Se muestra a continuación un cuadro que resume los resultados obtenidos luego de la aplicación del análisis factorial en los 5 cursos de extensión. Más adelante se detalla el proceso de obtención de los mismos de forma individual para cada curso.

RESULTADOS OBTENIDOS PARA LOS 5 CURSOS DE EXTENSIÓN UTILIZANDO ANALISIS FACTORIAL CON 5 FACTORES

Curso	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Explicación de la variación de los datos
Desarrollo de Habilidades Gerenciales	Desempeño	Infraestructura	Metodología	Contenido	Duración	68,20%
	<ul style="list-style-type: none"> •Calificación del taller (0,871) •Diferencia de conocimiento entre inicio y final (0,623) 	<ul style="list-style-type: none"> •Lugar del taller (0,932) 	<ul style="list-style-type: none"> •Ritmo y dinámica del taller (0,956) 	<ul style="list-style-type: none"> •Utilidad práctica del curso (0,769) •Temas tratados en el curso (0,614) 	<ul style="list-style-type: none"> •Cantidad de horas del taller (0,609) 	
Excel	Dispositivos didácticos	Desempeño	Duración/Organización	Desempeño	Infraestructura	85,80%
	<ul style="list-style-type: none"> •Claridad y exposición del docente (0,911) •Ritmo y dinámica del taller (0,763) •Motivación del participante (0,646) 	<ul style="list-style-type: none"> •Utilidad práctica del curso (0,903) •Calificación del taller en general (0,863) 	<ul style="list-style-type: none"> •Horario del taller (0,858) •Cantidad de horas del taller (0,719) 	<ul style="list-style-type: none"> •Diferencia de conocimiento entre inicio y final (0,956) 	<ul style="list-style-type: none"> •Lugar del taller (0,638) 	
Mercado de Capitales	Contenido	Dispositivos didácticos	Organización	Desempeño	-	72,60%
	<ul style="list-style-type: none"> •Utilidad práctica del curso (0,969) •Temas tratados en el curso (0,689) 	<ul style="list-style-type: none"> •Claridad y exposición del docente (0,864) •Motivación del participante (0,791) 	<ul style="list-style-type: none"> •Lugar del taller (0,879) •Horario del taller (0,770) 	<ul style="list-style-type: none"> •Calificación del taller en general (0,806) 		
Cadena de Suministros	Metodología	Organización	Desempeño	Pedagogía docente	Duración	83,90%
	<ul style="list-style-type: none"> •Ritmo y dinámica del taller (0,922) •Motivación del participante (0,805) 	<ul style="list-style-type: none"> •Lugar del taller (0,903) •Horario del taller (0,832) 	<ul style="list-style-type: none"> •Diferencia de conocimiento entre inicio y final (0,980) •Calificación del taller en general (0,610) 	<ul style="list-style-type: none"> •Claridad y exposición del docente (0,733) 	<ul style="list-style-type: none"> •Cantidad de horas del taller (0,819) 	
Metodologías Ágiles	Dispositivos didácticos	Organización	Organización	Grado de interés	-	84,50%
	<ul style="list-style-type: none"> •Claridad y exposición del docente (0,930) •Ritmo y dinámica del taller (0,908) •Temas tratados en el curso (0,834) 	<ul style="list-style-type: none"> •Horario del taller (0,890) •Diferencia de conocimiento entre inicio y final (0,826) 	<ul style="list-style-type: none"> •Lugar del taller (0,921) •Cantidad de horas del taller (0,633) •Calificación del taller en general (0,633) 	<ul style="list-style-type: none"> •Motivación del participante (0,988) 		



5- INTERPRETACIÓN DE LOS RESULTADOS

SS loadings: Son cargas factoriales (autovalores) que indican la importancia relativa de cada factor dando cuenta de la varianza asociada con el conjunto de variables que se analiza. La solución extrae los factores en orden de importancia, con el factor 1 con mayor varianza y el factor 5 con menor varianza. Estos valores se obtienen mediante la suma de los cuadrados de cada factor. Cada componente va explicando la proporción de variabilidad que no explican sus componentes anteriores.

Proportion var: Es la proporción de variabilidad en los datos explicada por cada factor. Valores más altos de porcentaje de la varianza indican que un factor explica una mayor parte de la variabilidad. Por lo tanto, se utilizan estos valores para determinar cuáles factores son más importantes.

Cumulative var: Representa la varianza acumulada explicada por la solución factorial. Luego de haber probado soluciones con 3, 4 y 5 factores para cada una de las bases de datos, se opta por utilizar este último dado que arroja el mayor porcentaje de variancia acumulada.

A partir de la salida obtenida por el software, se agrupan las variables que tienen influencias positivas más grandes (desde el 0,6 en adelante) en cada uno de los 5 factores correspondientes, pudiendo obtener las variables más importantes que destacaron en este curso.

5.1 Curso: Desarrollo de Habilidades Gerenciales

```
call:
factanal(x = DHG, factors = 5)

Uniquenesses:
  x1    x2    x3    x4    x5    x6    x7    x8    x9    x10
0.363 0.595 0.290 0.251 0.005 0.404 0.768 0.005 0.491 0.005

Loadings:
  Factor1 Factor2 Factor3 Factor4 Factor5
x1
x2    0.623
x3    0.419    0.303
x4    0.323    0.515    0.529    0.266    0.169
x5    0.213
x6
x7
x8
x9    0.337    0.466    0.329    0.261
x10   0.871    0.242    0.254

SS loadings      Factor1 Factor2 Factor3 Factor4 Factor5
Proportion Var   0.159   0.156   0.148   0.129   0.090
Cumulative Var   0.159   0.315   0.463   0.592   0.682

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 2.11 on 5 degrees of freedom.
The p-value is 0.834
>
```

Figura 6.1 Salida del software para el curso “DHG

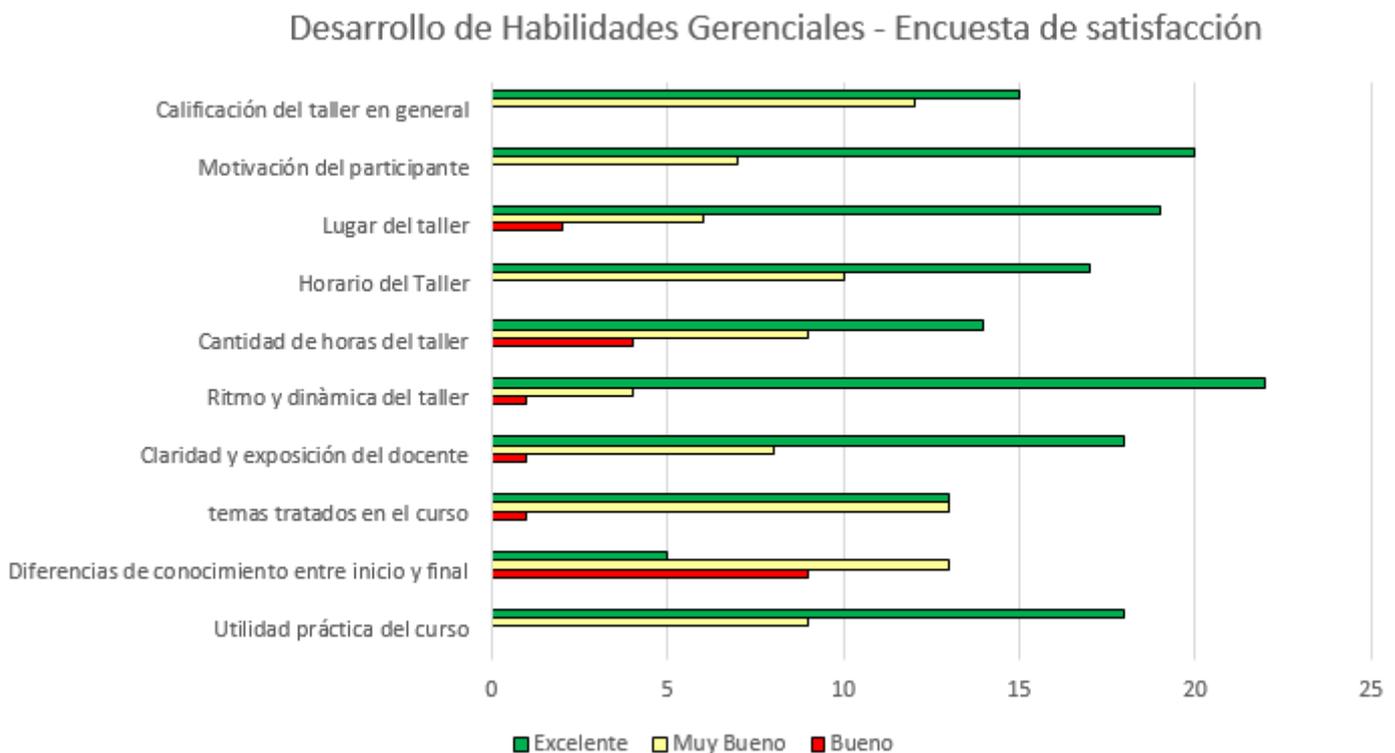


- Factor 1: Calificación del taller (0,871) y Diferencia de conocimiento entre inicio y final (0,623) tienen influencias positivas grandes en el Factor 1, por lo tanto este factor puede describir el desempeño del curso.
- Factor 2: Lugar del taller (0,932), este factor representa la infraestructura del curso.
- Factor 3: Ritmo y dinámica del taller (0,956), el factor 3 describe la metodología empleada en el curso.
- Factor 4: Utilidad práctica del curso (0,769) y Temas tratados en el curso (0,614) son variables de mayor influencia en el factor, por lo tanto este describe el contenido del taller.
- Factor 5: Cantidad de horas del taller (0.609), es la variable de mayor influencia positiva en este factor, describe la duración del curso.

En conjunto, los 5 factores explican un 68,2% de la variación de los datos. Es decir, el 68,2% del total de la varianza está representado por la información contenida en la matriz factorial para una solución de 5 factores.

De esta manera se logra condensar la información logrando una serie de variables más pequeña. A continuación se muestra un gráfico de barras que representa la manera en que fueron calificados cada uno de los atributos del curso, dejando en evidencia la correspondencia de estas calificaciones con las variables más importantes extraídas para cada factor mediante el empleo del análisis factorial:

Gráfico 5.1. Tipo de respuesta que dieron los participantes a cada variable





5.2 Curso: ExCel

Call:

```
factanal(x = EXC, factors = 5)
```

Uniquenesses:

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
	0.010	0.005	0.407	0.107	0.121	0.280	0.115	0.005	0.233	0.171

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
x1	0.113	0.903	0.175	0.318	-0.171
x2	0.157	0.156	0.206	0.946	
x3	0.288	0.566		0.169	0.401
x4	0.911	0.146	0.187		
x5	0.763	0.474	0.134		0.232
x6			0.719	0.428	0.131
x7	0.360	0.134	0.858		
x8	0.524	0.127	0.513	0.186	0.638
x9	0.646	0.177	0.226	0.390	0.340
x10	0.212	0.863			0.175

	Factor1	Factor2	Factor3	Factor4	Factor5
SS loadings	2.397	2.220	1.697	1.406	0.825
Proportion Var	0.240	0.222	0.170	0.141	0.083
Cumulative Var	0.240	0.462	0.631	0.772	0.855

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 4.7 on 5 degrees of freedom.
The p-value is 0.453

> |

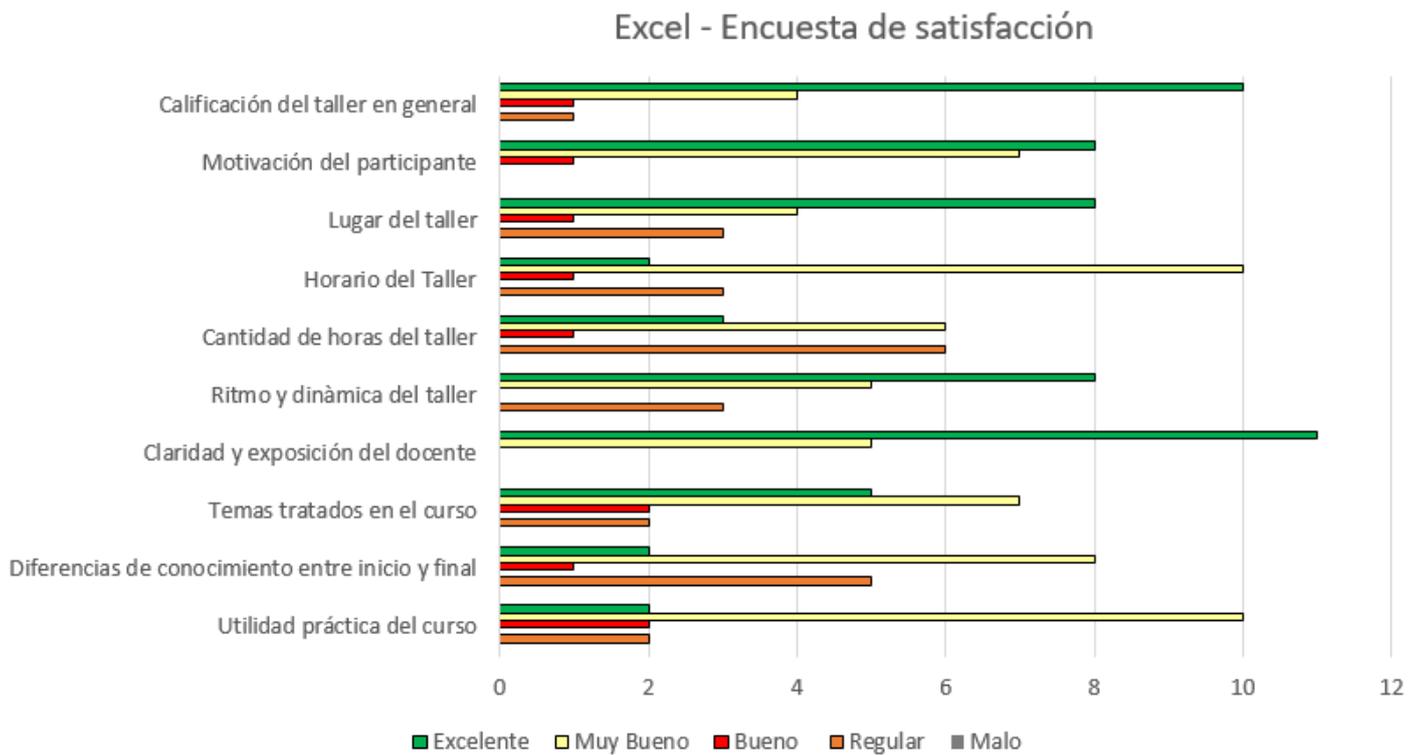
Figura 5.2 Salida del software para el curso "Excel"

- Factor 1: Claridad y exposición del docente (0,911), Ritmo y dinámica del taller (0,763) y Motivación del participante, son variables incluidas en este factor que representan los dispositivos didácticos del curso.
- Factor 2: Utilidad práctica del curso (0,903) y Calificación del taller en general (0,863), describen el desempeño del curso.
- Factor 3: Horario del taller (0,858) y Cantidad de horas del taller (0,719), representan la duración del taller.
- Factor 4: Diferencia de conocimiento entre inicio y final (0,956), este factor indica el desempeño del curso.
- Factor 5: Lugar del taller (0,638), indica la infraestructura del taller.

En conjunto, estos 5 factores explican un 85,8% de la variación de los datos. En la siguiente figura se muestra un gráfico que representa la forma en la que los alumnos de este curso calificaron las distintas variables:



Gráfico 5.2. Tipo de respuesta que dieron los participantes a cada variable





5.3 Curso: Mercado de Capitales

```
Call:
factanal(x = MC, factors = 5)

Uniquenesses:
      x1      x2      x3      x4      x5      x6      x7      x8      x9      x10
0.005 0.671 0.005 0.153 0.784 0.609 0.328 0.116 0.063 0.005

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
x1  0.969   0.163   0.172
x2  0.546           -0.136   0.109
x3  0.689   0.516  -0.132           0.482
x4  0.278   0.864           0.143
x5  0.350           0.286
x6  0.326  -0.343   0.186   0.325   0.163
x7           0.114   0.770           -0.237
x8           0.879   0.230   0.232
x9  -0.126   0.791   0.519   0.127
x10 0.105   0.387   0.424   0.806

      Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings  2.053   1.950   1.923   0.945   0.391
Proportion var 0.205   0.195   0.192   0.094   0.039
cumulative var 0.205   0.400   0.593   0.687   0.726

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 3.05 on 5 degrees of freedom.
The p-value is 0.693
> |
```

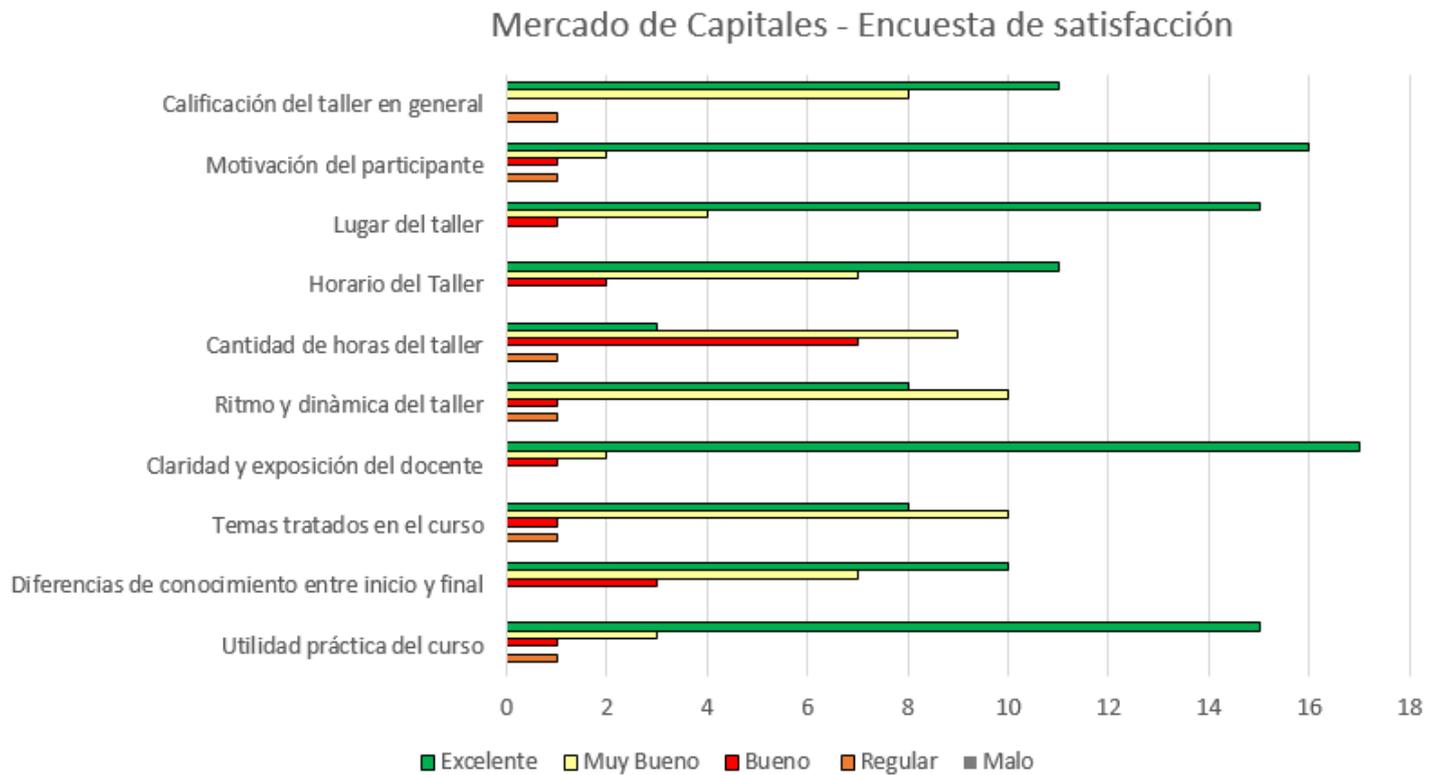
Figura 5.3 Salida del software para el curso “Mercado de capitales”

- Factor 1: Utilidad práctica del curso (0,969) y Temas tratados en el curso (0,689) son dos variables que componen este factor, el mismo representa el contenido del curso.
- Factor 2: Claridad y exposición del docente (0,864) y Motivación del participante (0,791), este factor describe los dispositivos didácticos del taller.
- Factor 3: Lugar del taller (0,879) y Horario del taller (0,770), el factor representa aspectos organizacionales.
- Factor 4: Calificación del taller en general (0,806), describe el desempeño del taller.
- Factor 5: No existen variables relevantes para este factor.

En conjunto, estos 5 factores describen un 72,6% de la variación de los datos. En la siguiente figura se muestra un gráfico que representa la forma en la que los alumnos de este curso calificaron las distintas variables:



Gráfico 5.3. Tipo de respuesta que dieron los participantes a cada variable





5.4 Curso: Cadena de Suministros

call:

```
factanal(x = cs, factors = 5)
```

Uniquenesses:

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
	0.530	0.005	0.488	0.078	0.005	0.086	0.034	0.005	0.124	0.258

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5
x1			0.235	-0.639	
x2	0.121		0.980	-0.112	
x3	0.165		0.369	-0.373	-0.452
x4	0.443	0.333	0.250	0.733	0.127
x5	0.922	0.271		0.118	0.237
x6	0.317	0.319	0.139	-0.150	0.819
x7		0.732	0.115	0.382	0.514
x8	0.389	0.903	0.101	0.106	
x9	0.805		0.447	0.125	
x10	0.212	0.546	0.610	-0.133	0.101

	Factor1	Factor2	Factor3	Factor4	Factor5
SS loadings	2.043	1.954	1.829	1.324	1.238
Proportion Var	0.204	0.195	0.183	0.132	0.124
Cumulative Var	0.204	0.400	0.583	0.715	0.839

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 3.94 on 5 degrees of freedom.
The p-value is 0.558

> |

Figura 5.4 Salida del software para el curso “Cadena de suministros”

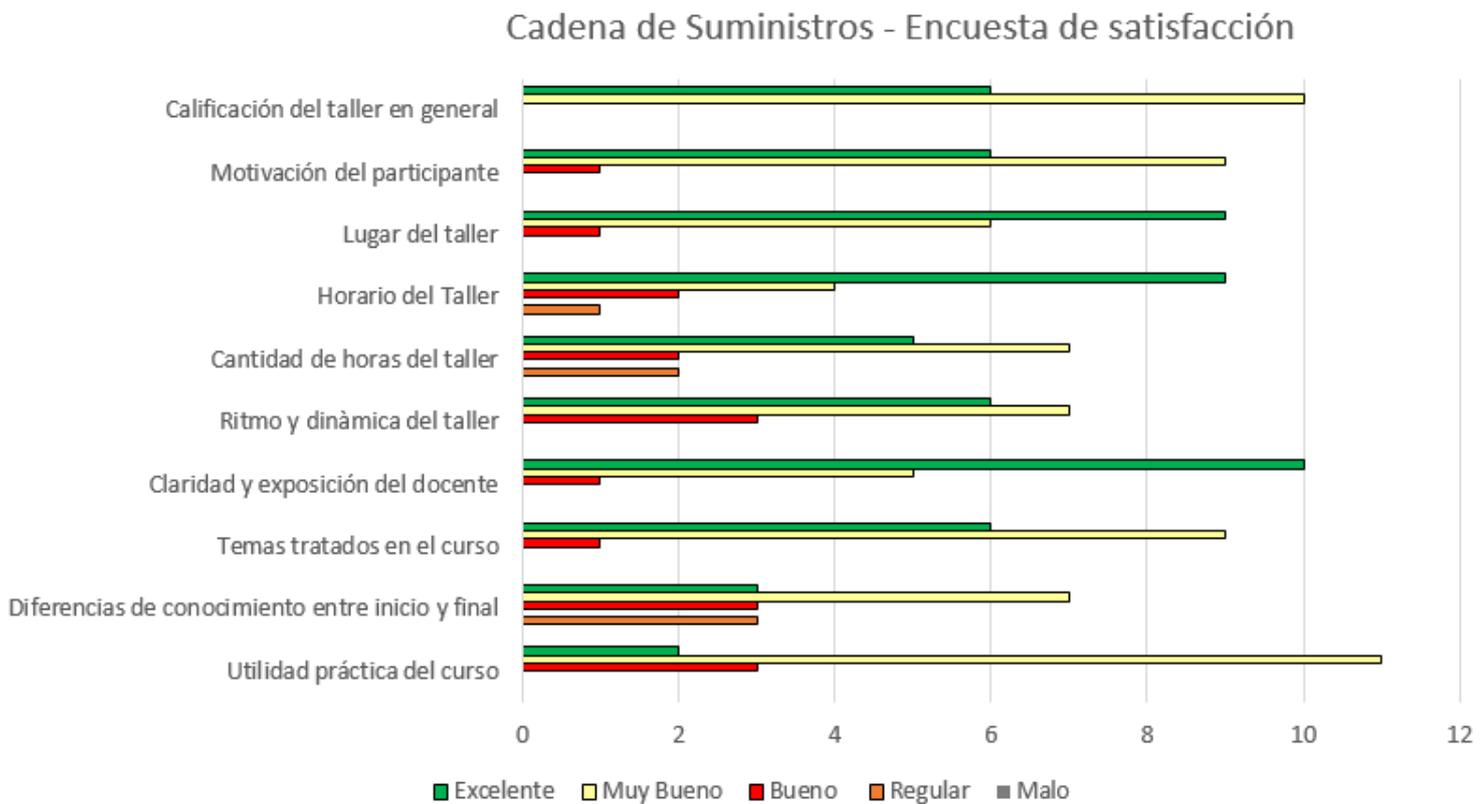
- Factor 1: Ritmo y dinámica del taller (0,922) y Motivación del participante (0,805) son las variables con influencias positivas más grandes en este factor que representa la metodología del curso.
- Factor 2: Lugar del taller (0,903) y Horario del taller (0,832), este factor describe la organización del taller.
- Factor 3: Diferencia de conocimiento entre inicio y final (0,980) y Calificación del taller en general (0,610) son las variables más importantes en este factor que describe el desempeño del curso.
- Factor 4: Claridad y exposición del docente (0,733), este factor describe la pedagogía docente.
- Factor 5: Cantidad de horas del taller (0,819), es un factor que describe la duración del taller.

En conjunto, estos 5 factores describen un 83,9% de la variación de los datos.



En la siguiente figura se presenta un gráfico de barras que muestra la manera en la que fueron calificadas las distintas variables del curso:

Gráfico 5.4. Tipo de respuesta que dieron los participantes a cada variable





5.5 Curso: Metodologías Ágiles

```
Call:  
factanal(x = MA, factors = 5)
```

```
Uniquenesses:  
  x1    x2    x3    x4    x5    x6    x7    x8    x9    x10  
0.260 0.291 0.177 0.015 0.126 0.544 0.005 0.096 0.005 0.030
```

```
Loadings:  
      Factor1 Factor2 Factor3 Factor4 Factor5  
X1  -0.171   0.572   0.155           0.597  
X2           0.826           0.101  
X3   0.834   0.217   0.277           0.101  
X4   0.930  -0.150           -0.294  
X5   0.908  -0.110           0.191  
X6   0.111           0.633  -0.192  
X7  -0.127   0.890   0.413  -0.113  
X8           0.220   0.921  
X9           -0.107   0.988  
X10  0.309  -0.126   0.633   0.401   0.545
```

```
          Factor1 Factor2 Factor3 Factor4 Factor5  
SS loadings  2.552  1.951  1.951  1.195  0.801  
Proportion Var  0.255  0.195  0.195  0.120  0.080  
Cumulative Var  0.255  0.450  0.645  0.765  0.845
```

```
Test of the hypothesis that 5 factors are sufficient.  
The chi square statistic is 2.9 on 5 degrees of freedom.  
The p-value is 0.716  
> |
```

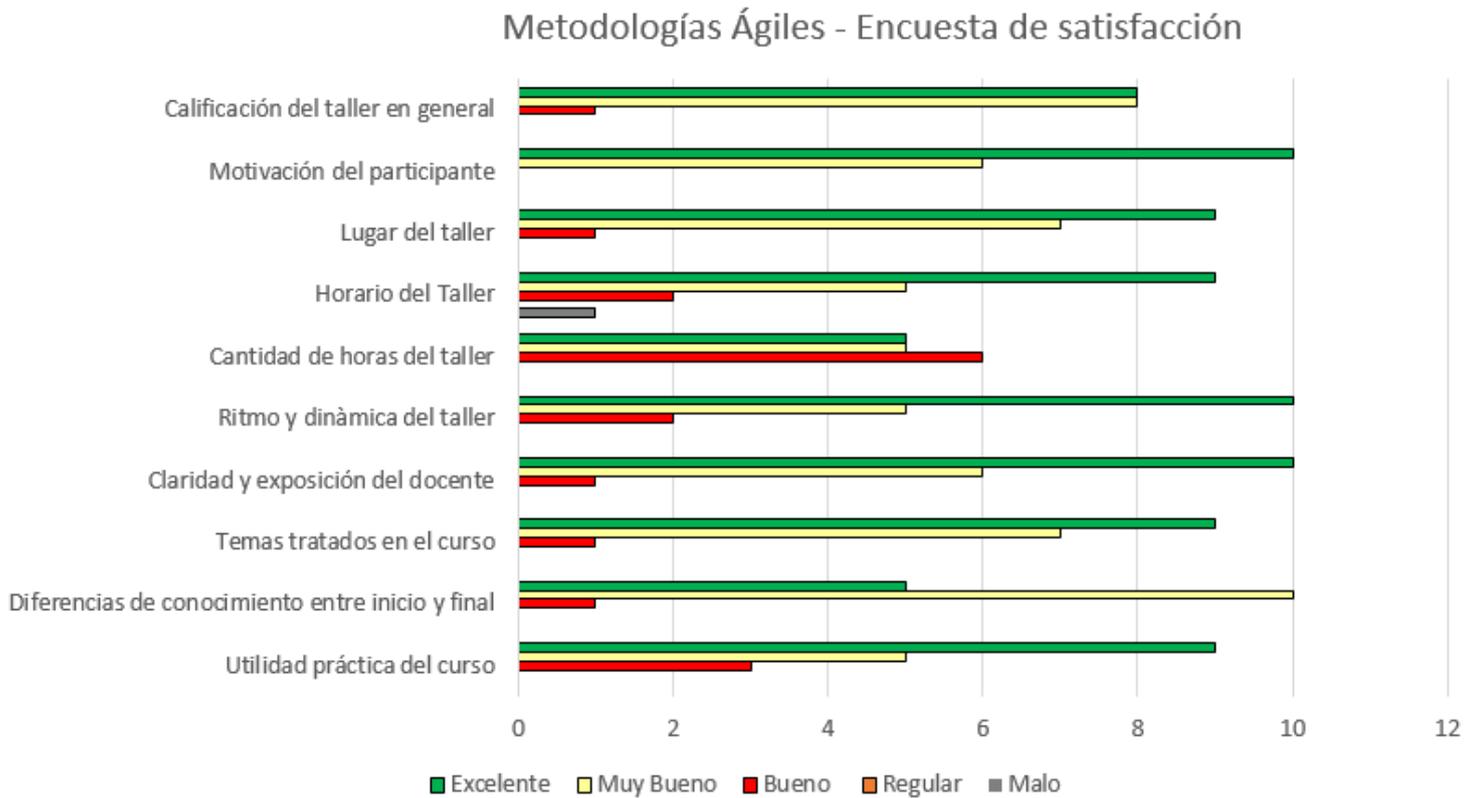
Figura 5.5. Salida del software para el curso “Metodologías ágiles”

- Factor 1: Claridad y exposición del docente (0,930), Ritmo y dinámica del taller (0,908) y Temas tratados en el curso (0,834), son los atributos que más importancia tuvieron para este factor, el cual representa los dispositivos didácticos del curso.
- Factor 2: Horario del taller (0,890) y Diferencia de conocimiento entre inicio y final (0,826), este factor describe la organización del taller.
- Factor 3: Lugar del taller (0,921), Cantidad de horas del taller (0,633) y Calificación del taller en general (0,633), son variables que agrupadas en este factor representan la organización del taller.
- Factor 4: Motivación del participante (0,988), este factor describe el grado de interés del alumno.
- Factor 5: No existen variables relevantes para este factor.

En conjunto, los 5 factores representan un 84,5% de la variación de los datos. En la siguiente figura se representa la manera en la que fueron calificadas cada una de las variables del curso:



Gráfico 5.5. Tipo de respuesta que dieron los participantes a cada variable





6- CONCLUSIONES

Teniendo una visión global de los resultados obtenidos mediante un análisis factorial para cada uno de los cursos de extensión, se puede observar cuales son los aspectos de un curso que los usuarios más valoran.

Luego de haber condensado las distintas variables en una serie de factores, se muestra aquellos que más destacaron ordenados según la cantidad de veces que se repiten a lo largo de cada uno de los cursos, estos son:

- Desempeño
- Organización
- Dispositivos didácticos
- Infraestructura
- Metodología
- Contenido
- Duración
- Pedagogía docente
- Grado de interés

La finalidad de esta aplicación es comprender más fácilmente lo que los usuarios opinan del curso y cuáles son las variables que consideran más importantes. Otro aspecto fundamental a tener en cuenta luego de estos resultados, es que al momento de direccionar la oferta de un taller de capacitación se pueden tener en consideración los distintos factores que abarcan los aspectos que los usuarios más aprecian, para seguir manteniéndolos y reforzándolos en cada futuro curso de extensión y para mejorar aquellas variables más débiles que fueron juzgadas con puntaje más bajo.

Al momento de analizar las soluciones obtenidas resulta indispensable pensar de qué forma se podría despertar el interés de las personas con el fin de que puedan ampliar su base de conocimiento, mejorar su crecimiento profesional y personal, los cuales son la finalidad de todo curso de formación.

La Facultad de Ciencias Económicas podría mejorar aún más su prestigio y aumentar su alumnado al poner foco en su oferta de programas de actualización, formación y capacitación como lo son sus cursos de extensión, ofrecidos por el Instituto de Administración.

Habiendo trabajado con Big Data y la analítica de negocios, y dados los resultados del análisis factorial se puede asegurar que los aspectos que requieren mayor esfuerzo por mejorar a futuro son los relacionados con la pedagogía docente y el grado de interés del alumno.

Con respecto a la “pedagogía docente” parecería necesario comprender la base que debe tener un formador de conocimiento con respecto al proceso educativo y a la formación de un sujeto, se evidencia poco considerada la variable “Claridad y



exposición del docente” por los alumnos. Se debería dejar de lado el aprendizaje reproductivo donde los alumnos repiten los conocimientos aportados por el docente y en lugar de ello proyectar un aprendizaje creador que permitirá a los alumnos generar su propia iniciativa y desenvolverse en un ambiente problemático con su propia creatividad. Asimismo se podrá enriquecer la utilidad práctica del curso y su calificación general final.

Este aspecto permitiría a su vez, mejorar el “grado de interés” que incluye la motivación del participante, que es otra variable con baja relevancia en los resultados obtenidos por el análisis factorial. Como se mencionó, el aprendizaje obtenido por medio de la repetición, la copia y reproducción de los conocimientos aportados por el docente, hoy en día, no parecen motivar al alumnado.

En el otro extremo, se encuentran los factores más relevantes que son el “desempeño” y la “organización” de los cursos. Siendo estos los aspectos que mejor fueron tenidos en cuenta por los participantes, parecería indiscutible el saber que son cualidades que no se deben perder de vista. Reforzar estos elementos podría implicar en un futuro la oferta de acceso a programas de capacitación mediante cursos E-learning o cursos con modalidad intensiva que permitirá a la facultad ampliar su alumnado, dado que hay personas que podrían optar por no tomar un curso presencial simplemente porque el mismo se extiende en el tiempo o el lugar u horario no les resultan convenientes.



1- BIBLIOGRAFÍA

- J. F. Hair, Jr., R. E. Anderson, R. L. Tatham, W. C. Black (1999). Análisis Multivariante, 5° edición. Iberia, Madrid. Prentice Hall.
- Brian S. Everitt and Torsten Hothorn (2005). A Handbook of Statistical Analyses Using R. Recuperado de <https://static1.squarespace.com/static/57af6d69b3db2b8d0eade247/t/580a6f849f74564489f129dc/1477078927168/R2.pdf>
- Barry Render, Ralph Stair, Michael Hanna, Trevor Hale (2016). Métodos cuantitativos para los negocios, 12° edición. México. Pearson.
- Daniel Peña (2002). Análisis multivariante de datos. Mc Graw Hill Interamericana
- Del bit... al Big Data. Recuperado de http://cdn2.hubspot.net/hub/239039/file-359994269-pdf/docs/PowerData_-_Del_bit%E2%80%A6_Al_Big_Data.pdf



Universidad Nacional de Tucumán
Facultad de Ciencias Económicas
Instituto de Administración
Práctica Profesional



ANEXOS



Anexo I: Modelo de encuesta de satisfacción empleada en los cursos de extensión.

 Universidad Nacional de Tucumán
 "100 años iluminando el pasado, 100 años proyectando el futuro"
 Facultad de Ciencias Económicas
 Instituto de Administración



ENCUESTA CURSO DE EXTENSIÓN

CURSO DE EXTENSIÓN: Hotelería y Asesoría

Edad: 23 Sexo F M Carrera: CPN, ... LAE LEE, ... ¿Trabajas? SI NO

Situación: Graduado Estudiante Cantidad de Materias Aprobadas

Marque con una cruz (X) en el casillero correspondiente en la escala de 1 a 5 donde:
 1 = Malo; 2 = Regular; 3 = Bueno; 4 = Muy Bueno; 5 = Excelente

	1	2	3	4	5
Utilidad práctica del curso				<input checked="" type="checkbox"/>	
Diferencias de conocimiento entre inicio y final					<input checked="" type="checkbox"/>
Temas tratados en el Curso					<input checked="" type="checkbox"/>
Claridad y Exposición del Docente					<input checked="" type="checkbox"/>
Ritmo y dinámica del taller					<input checked="" type="checkbox"/>
Cantidad de horas del taller					<input checked="" type="checkbox"/>
Horario del taller					<input checked="" type="checkbox"/>
Lugar del taller					<input checked="" type="checkbox"/>
Motivación del participante					<input checked="" type="checkbox"/>
Calificación del taller en general					<input checked="" type="checkbox"/>

Aspectos Positivos del Curso	Aspectos a Mejorar del Curso

¿Recomendaría realizar este curso? ¿Por qué? SI