

TRABAJO DE GRADO
Opción Seminario-Diplomado.

Satisfacción del Cliente en la aerolínea Delta: Factores Claves Mediante el Análisis de Datos

Corporación Universitaria Remington
Facultad de Ingenierías
Ingeniería de Sistemas

Carolina Alzate Sanchez¹
Katherine Carvajal Castrillon²
Ivonne Castaño Osorio³
John Edison Amortegui Granada⁴
Seminario-Diplomado
2024

¹ Estudiante de séptimo semestre de Ingeniería de Sistemas, Uniremington sede Pereira, Risaralda E. mail.: carolina.alzate.1731@miremington.edu.co

² Estudiante de séptimo semestre de Ingeniería de Sistemas, Uniremington sede Pereira, Risaralda E. mail.: angie.carvajal.0771@miremington.edu.co

³ Tutor temático seminario de Big Data y Ciencia de Datos. E-mail.: ivonne.castano@uniremington.edu.co

⁴ Tutor metodológico seminario Big Data y Ciencia de Datos. E-mail.: john.amortegui@uniremington.edu.co

Tabla de Contenidos

Resumen.....	4
Palabras clave: Análisis de Datos, Crisp-dm, Visualización, Satisfacción del Cliente, Big Data..	4
Pregunta orientadora de la búsqueda	5
Metodología de búsqueda de la información:	6
Entendimiento del negocio:	6
Enfoque analítico:	6
Requisitos de datos:	7
Recopilación de datos:	7
Comprensión de datos:.....	7
Preparacion de datos:	8
Implementación:	8
Retroalimentación:.....	8
Sustentación teórica de la pregunta.....	9
Modelado:	9
Evaluación del modelo:.....	10
Servicio a bordo:	11
Comodidad del asiento:.....	12
Conclusiones.	14
Lista de referencias	15

Lista de tablas

Tabla 1	10
Tabla 2	11
Tabla 3	12
Tabla 4	13

Resumen

El propósito de este texto es enfocarse principalmente en la satisfacción del cliente con los servicios de las aerolíneas aplicando la metodología CRIPS-DM. Dicha metodología, se utiliza en sus diez fases principales, que se tendrán presentes para desarrollar el proceso de análisis como lo son: la comprensión del negocio, la preparación de los datos, el modelado y la evaluación que ayudarán a la resolución de los factores determinantes en satisfacción de los clientes. De igual manera, es preciso completar cada una de las fases para continuar con la siguiente y obtener un buen resultado. El proceso incluye la verificación de la disposición de los datos necesarios, la clasificación y la modificación de los datos a los formatos idóneos para el análisis y el uso de técnicas de modelado junto con la verificación de que los modelos cumplan con los objetivos establecidos, utilizando un enfoque correlacional. En relación con el análisis de este conjunto de datos, se alcanza a obtener información valiosa para las aerolíneas de los factores que contribuyen a la satisfacción del cliente, permitiendo la adaptabilidad de sus servicios a causa de lo que antes se ha dicho para mejorar la experiencia general del cliente.

Palabras clave: Análisis de Datos, Crisp-dm, Visualización, Satisfacción del Cliente, Big Data.

Pregunta orientadora de la búsqueda

El nivel de satisfacción de los clientes permite a las empresas tomar decisiones sobre cuales servicios mejorar e invertir continuamente. En la compañía aérea Delta se evalúan varios factores con respecto a la satisfacción de los usuarios según los servicios prestados, e interesa conocer qué es lo realmente importante para ellos utilizando metodologías relacionadas con el Análisis de Datos. Según lo anterior surge la siguiente pregunta:

¿Cuáles son los factores claves que determinan la satisfacción en el cliente de la compañía aérea Delta?

Metodología de búsqueda de la información:

Se implementa en el proyecto de análisis la metodología CRISP-DM, fue creada a finales de 1996 y se posicionó como una metodología organizada y a su vez es un estándar para llevar a cabo proyectos de extracción de datos.

La metodología CRISP-DM proporciona una guía estructurada y detallada para llevar a cabo proyectos de minería de datos, dividiendo el proceso en seis fases clave: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. (Chapman et al., 2000, p. 3)

Entendimiento del negocio:

En esta primera fase, el objetivo es entender las metas y requerimientos del problema que se requiere resolver. “Se establece cuáles serán los criterios para medir el éxito en el proyecto, ya sean de tipo cualitativo o cuantitativo.” (IPMOGuide, s.f). Es importante comprender con claridad los objetivos, los recursos disponibles y las metas específicas al hacer el análisis de los datos.

Enfoque analítico:

Se emplea el estudio descriptivo y correlacional, se enfoca en responder a preguntas sobre el qué, el dónde y el cuándo al momento de sintetizar los datos y sin constituir relaciones entre estas (CIMEC, 2023). El descriptivo es donde se recopilan datos sin alterar deliberadamente las variables o condiciones.

En primer lugar, la identificación de los patrones hace más fácil la interpretación detallada de un fenómeno o que influyen en una determinada población, mientras, se comprenden mejor los comportamientos y actitudes.

Requisitos de datos:

Dentro de los principales requisitos es necesario disponer de variables que tengan relación con el tema a evaluar y cada variable debe tener una calificación correspondiente, para poder estudiar los datos y obtener el resultado deseado, asimismo, se requiere una base de datos en tipo csv.

Recopilación de datos:

Con respecto a la base de datos necesaria para trabajar, se plantea descargar desde la página kaggle un archivo en formato csv, donde se elige un tema en específico como la satisfacción del cliente en una aerolínea, inicialmente identificar que no hay espacios vacíos ni textos nulos con la intención de analizar los datos que permitan dar respuesta a la pregunta especificada en el documento.

Comprensión de datos:

La fase de comprensión de los datos empieza con la recopilación inicial de datos y avanza con actividades destinadas a familiarizarse con ellos, se identifican los problemas de calidad y ayuda a detectar subconjuntos interesantes para generar hipótesis sobre información oculta, se procura tener un conocimiento detallado de la información disponible para transformar las variables (Chapman et al., 2000, p. 10).

En este mismo sentido Maldonado (2022) afirma que:

Diversos artículos y estudios de parte de las organizaciones como IBM quienes mencionan que más de 150 zettabytes (150 billones de gigas) requieren de distintos análisis para el 2025 o Forbes quienes descubrieron que se crean 2,5 trillones de bytes de datos cada día. (p. 20)

Preparación de datos:

La preparación de los datos es una de las tareas más cruciales dentro del proceso de minería de datos, ya que suele llevar entre el 50% y el 70% del tiempo y esfuerzo de un proyecto. Dependiendo de la organización y sus objetivos, la preparación de datos típicamente involucra tareas como fusionar conjuntos de datos y/o registros, la selección de un subconjunto de datos de muestra, ordenar los datos para el modelado y eliminar o reemplazar valores en blanco o faltantes. (IBM Corporation, 2021)

Implementación:

La fase de implementación depende de la evaluación final del modelo utilizado para el proyecto de análisis,

Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso. (Nay Lee, 2016)

Para este proyecto de análisis de datos no se realiza la fase de implementación, ya que esta se ve más en el análisis de modelos predictivos y el presente proyecto se enfoca en el análisis descriptivo y correlacional.

Retroalimentación:

En esta fase, después de haber terminado con todo el proceso de análisis se identifican los aspectos a mejorar en el proceso de modelado y que posiblemente

necesitan de un análisis diferente o indagar más en los datos para poder obtener una respuesta concreta y realizar ajustes necesarios basándonos en las observaciones realizadas por los usuario o expertos.

No se lleva a cabo el proceso de retroalimentación en este análisis por el tiempo limitado del proceso de modelado y ejecución final del proyecto.

Sustentación teórica de la pregunta

Modelado:

En el proceso de modelado se incorporan las herramientas de análisis donde se integran los datos relevantes que darán respuesta la pregunta planteada al inicio del proyecto. “Normalmente, los mineros de datos ejecutan varios modelos utilizando los parámetros predeterminados y luego ajustan los parámetros o vuelven a la fase de preparación de datos para las manipulaciones requeridas por el modelo de su elección”. (IBM Corporation, 2021)

Una de las herramientas a utilizar es Excel-Powequery que permite la transformación de datos, eliminación de variables no relevantes y el cambio del tipo de dato para un mejor tratamiento de los datos. También se implementa la técnica de regresión, “el análisis de regresión describe la dependencia causal entre la variable dependiente (y) y la variable independiente (x) para predecir valores y en términos de x”. (Fernández, 2013, p. 17)

Se aplica el coeficiente de correlación para determinar la relación entre dos propiedades.

Tabla 1

Regresión de variables de análisis Delta

Resumen									
<i>estadísticas de la regresión</i>									
Coefficiente	0,729424								
Coefficiente	0,53206								
F ² ajustac	0,531986								
Error típico	0,418072								
Observacio	89998								
ANÁLISIS DE VARIANZA									
	<i>Grados de libertad</i>	<i>de cuadrados</i>	<i>de los cu.</i>	<i>F</i>	<i>valor crítico de F</i>				
Regresión	13	17883,06	1375,62	7870,395	0				
Residuos	89985	15727,94	0,174784						
Total	89998	33611							
<i>Coefficientes Error típico Estadístico t Probabilidad superior 95% inferior 95% superior 97,0% inferior 97,0%</i>									
Intercepció	0	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
265	-5,8E-05	1,39E-06	-41,7239	0	-6,1E-05	-5,5E-05	-6,1E-05	-5,5E-05	
0	0,035592	0,001641	21,68448	5,3E-104	0,032375	0,038809	0,03203	0,039154	
0	-0,05705	0,001575	-36,2308	2,3E-285	-0,06014	-0,05396	-0,06047	-0,05363	
2	-0,01139	0,001154	-9,86969	5,79E-23	-0,01366	-0,00913	-0,0139	-0,00889	
2	-0,03567	0,001607	-22,2036	6,2E-109	-0,03882	-0,03252	-0,03916	-0,03219	
4	0,130055	0,001365	95,27807	0	0,127379	0,13273	0,127092	0,133017	
2	0,01363	0,001587	8,588435	8,95E-18	0,010519	0,01674	0,010186	0,017074	
3	0,045253	0,002031	22,28046	1,1E-109	0,041272	0,049234	0,040845	0,049661	
3	0,034255	0,001348	25,40904	6,4E-142	0,031613	0,036898	0,03133	0,037181	
3	-0,00991	0,001565	-6,33086	2,45E-10	-0,01298	-0,00684	-0,01331	-0,00651	
5	0,029183	0,00118	24,73192	1,4E-134	0,02687	0,031496	0,026622	0,031744	
3	-0,01237	0,001608	-7,68907	1,5E-14	-0,01552	-0,00921	-0,01585	-0,00888	
2	0,000988	0,001907	0,518123	0,604374	-0,00275	0,004725	-0,00315	0,005126	

Nota. se muestra la regresión de la totalidad de las variables a analizar de la aerolínea

Delta. Fuente: Elaboración propia.

Evaluación del modelo:

La evaluación permite no solo evaluar la calidad del modelo, sino también actuar como punto de control para verificar si cumple con la solicitud original. Este paso asegura que el modelo no solo tenga un buen rendimiento en cuanto a precisión y generalización, sino que también se ajuste a las necesidades y metas específicas mencionadas al inicio del proyecto. Los datos en la regresión total realizada con las variables de los servicios prestados por la aerolínea Delta, representan valores vinculados

a fenómenos bajo investigación, mostrando las conexiones entre variables dependientes e independientes. Se calculan las relaciones entre ellas y se utilizan técnicas de regresión para estimar los parámetros del modelo. (Fernández, 2013)

Los factores relevantes para la satisfacción del cliente de la aerolínea Delta según el análisis de regresión son: Servicio a bordo, el embarque en línea, el servicio de check-in y la facilidad de reserva en línea.

Los factores menos relevantes para la satisfacción del cliente de la aerolínea Delta según el análisis de regresión son: Comodidad del asiento, la ubicación de la puerta y el servicio de espacio para las piernas.

Tabla 2

Correlación de variables de análisis Delta

	Flight diagnostico	Distance Distanc	Seat Seat	Comfort comfort	Food and	Drink drink	Gate locatio	Inflight wifi	Inflight entertai	Online Online	Ease of bookin	On- board	Baggag e	Checkin service	Cleanlin ess	Online boardin
diagnostico	1															
Flight Distance	-0,124	1														
Seat comfort	0,228	-0,053	1													
Food and drink	0,1068	-0,017	0,7338	1												
Gate location	-0,023	-0,003	0,2179	0,3755	1											
Inflight wifi service	0,1555	-0,087	0,0963	0,0095	-0,005	1										
Inflight entertainment	0,4395	-0,133	0,5144	0,4715	-0,003	0,1746	1									
Online support	0,276	-0,133	0,0767	-0,005	0,0005	0,6082	0,3358	1								
Ease of Online booking	0,2421	-0,072	0,1566	-0,004	-0,002	0,6835	0,1699	0,6102	1							
On-board service	0,23	-0,025	0,1281	0,0137	-0,037	0,0491	0,0939	0,0758	0,3357	1						
Baggage handling	0,1426	0,0344	0,1405	0,0169	-9E-04	0,0227	0,0243	0,0184	0,319	0,4785	1					
Checkin service	0,2365	-0,039	0,0658	0,0005	-0,043	0,0787	0,1846	0,165	0,0969	0,244	0,2287	1				
Cleanliness	0,1443	0,0255	0,1299	0,0124	-0,004	0,0256	0,0185	0,0145	0,3388	0,5028	0,6209	0,2352	1			
Online boarding	0,237	-0,084	0,0942	-0,018	-0,004	0,7101	0,2533	0,699	0,7368	0,0979	0,0719	0,164	0,0684	1		

Nota. se muestra la correlación de las variables en la aerolínea Delta. Fuente: Elaboración propia.

Servicio a bordo:

El modelo de regresión con un R^2 ajustado de 0.94 y un valor p de 0.003 es muy robusto y fiable, indicando que el servicio a bordo está muy bien explicado por las

variables del modelo, y que estas variables tienen una relación significativa y real con el nivel de servicio a bordo. El resultado es altamente significativo, en otras palabras, hay fuertes evidencias contra la hipótesis nula.

Tabla 3.

Regresión de servicio abordó.

SERVICIO A BORDO

Estadísticas de la regresión

Coeficiente t	0,97803327
Coeficiente r	0,95654907
R ² ajustado	0,94206542
Error típico	0,22034088
Observación	5

ANÁLISIS DE VARIANZA

		Grados de libertad	de cuadrado de los cuadrados	F	valor crítico de F
Regresión	1	3,20641374	3,20641374	66,0433988	0,00389534
Residuos	3	0,14565031	0,0485501		
Total	4	3,35206404			

	Coeficientes	Error típico	Estadístico t	Probabilidad Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%
Intercepción	-1,74813739	0,23109546	-7,56456827	0,00479125	-2,48358628	-1,01268849	-2,48358628
SERVICIO A BORDO	0,56625204	0,0696779	8,12670898	0,00389534	0,34450586	0,78799823	0,34450586

Nota. se muestra la regresión de la variable de servicio a bordo de la aerolínea Delta.

Fuente: Elaboración propia.

Comodidad del asiento:

El r² ajustado de -0.24 indica que el modelo actual es ineficiente para poder predecir la probabilidad de analizar la satisfacción del cliente. Hay una relación significativa con la probabilidad del 0.88, la comodidad del asiento no tiene efecto sobre la satisfacción del cliente, es decir, la satisfacción es la misma sin importar la selección de asientos, por lo tanto, es una hipótesis nula.

Tabla 4.

Regresión de comodidad del asiento.

COMODIDA DEL ASIENTOEstadísticas de la regresión

Coefficiente (0,0753922
Coefficiente (0,00568398
R^2 ajustad	-0,24289502
Error típico	3,34571013
Observacion	6

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Media cuadrada	F	Valor crítico de F
Regresión	1	0,25595579	0,25595579	0,0228659	0,88712597
Residuos	4	44,7751052	11,1937763		
Total	5	45,031061			

	Coefficientes	Error típico	Estadístico t	Probabilidad Inferior 95%	Superior 95%	Inferior 95,0%	Superior 95,0%	
Intercepción	2,01129557	2,42144722	0,83061714	0,4528917	-4,71171971	8,73431085	-4,71171971	8,73431085
COMODIDA I	-0,12093819	0,79977769	-0,15121475	0,88712597	-2,34147705	2,09960068	-2,34147705	2,09960068

Nota. se muestra la regresión de la variable de Comodidad del asiento de la aerolínea

Delta. Fuente: Elaboración propia.

Conclusiones.

En el proceso de ciencia de datos que se ha realizado para este proyecto, se puede evidenciar que es bastante importante el estudio de los datos y las relaciones que puedan tener entre ellos para poder dar buenas respuestas para las preguntas que surgen con los datos generados, ya que más allá de simplemente organizar los datos, entenderlos y analizarlos cuando se tiene una buena visualización se puede entender mucho más fácil lo que dichos datos quieren decir.

Realizando este proyecto podemos ver que es bastante notorio que la metodología CRISP-DM que fue la utilizada en esta ocasión es bastante útil, ya que no da un paso a paso donde se puede complementar cada paso con el siguiente para obtener los resultados deseados, se hizo una correcta preparación y se pueden realizar los análisis correspondientes con el uso de herramientas que nos ayudan para poder realizar un análisis completo.

Lista de referencias

- IBM Corporation. (2021). *Manual CRISP-DM*. North Castle Drive: IBM SPSS Modeler.
- Chapman, P. C. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Tübingen: SPSS Inc.
- CIMEC. (14 de Julio de 2023). *CIMEC*. Obtenido de <https://www.cimec.es/investigacion-descriptiva-caracteristicas/>
- HUSEYN, R. (s.f). *Kaggle*. Obtenido de <https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction>
- IPMOGuide. (s.f). *IPMOGuide*. Obtenido de <https://ipmoguide.com/crisp-dm-metodologia-de-datos/>
- NAY LEE MOJARRANGO, J. E. (2016). unach. *ANÁLISIS COMPARATIVO DE LAS PLATAFORMAS WEKA Y MICROSOFT* . Ecuador. Obtenido de <http://dspace.unach.edu.ec/bitstream/51000/1624/1/UNACH-EC-ISC-2016-0007.pdf>
- Pat Fernández, L. A. (2013). *Introducción a los modelos de regresión: (ed.)*. Mexico: Plaza y Valdés. Obtenido de <https://elibro.net/es/lc/remington/titulos/39173>
- Sebastián Maldonado, C. V. (2022). *Analytics y Big Data: ciencia de los Datos aplicada al mundo de los negocios: (1 ed.)*. Santiago: RIL editores.
- Stewart, L. (s.f). *ATLAS.ti*. Obtenido de <https://atlasti.com/es/research-hub/investigacion-descriptiva>