

Proyecto Integrador

Hallazgos y análisis dentro de una organización

Grupo 4_Trabajo_Final_Data_Analytics1_FOLCADEMY

Grupo 4

Alegre Colombo, María de los Ángeles

González, Valeria Vanesa

Mayol, Hernán

Rodríguez Fittabile, Florencia

Vázquez, Santiago

Entrega 1: Análisis Inicial, Limpieza y Planteamiento de Hipótesis

Definición del Problema:

-Identificar las **causas** más comunes de interrupciones en los vuelos mediante el estudio de la frecuencia de **cancelaciones** por parte de las **aerolíneas** en un día particular del año 2015.



Hipótesis

La eficiencia de una aerolínea está correlacionados positivamente con la baja incidencia de cancelaciones de vuelos. A medida que la frecuencia de cancelaciones de vuelos aumenta en diferentes regiones, la calidad del servicio ofrecido por la empresa tiende a disminuir, lo que puede afectar negativamente su desempeño y reputación en el mercado.



Limpieza de Datos en archivo Excel

Trabajo Final_Grupo 4 _Vuelos.xlsx

Trabajo Final_Grupo 4 _Vuelos .XLSX

Archivo Editar Ver Insertar Formato Datos Herramientas Ayuda

110% \$ % .0 .00 123

F6 | fx

	A	B
1	AIRLINE_CODE	AIRLINE_NAME
2	UA	United Air Lines Inc.
3	AA	American Airlines Inc.
4	US	US Airways Inc.
5	F9	Frontier Airlines Inc.
6	B6	JetBlue Airways
7	OO	Skywest Airlines Inc.
8	AS	Alaska Airlines Inc.
9	NK	Spirit Air Lines
10	WN	Southwest Airlines Co.
11	DL	Delta Air Lines Inc.
12	EV	Atlantic Southeast Airlines
13	HA	Hawaiian Airlines Inc.
14	MQ	American Eagle Airlines Inc.
15	VX	Virgin America
16		

+ ≡ airline flight airport cancellation

Trabajo Final_Grupo 4 _Vuelos .XLSX

Archivo Editar Ver Insertar Formato Datos Herramientas Ayuda

110% \$ % .0 .00 123 Arial 10 B I A

G34 | fx

	A	B	C
1	IATA_CODE_AIRPORT	AIRPORT_NAME	CITY
2	ABR	Aberdeen Regional Airport	Aberdeen
3	ABI	Abilene Regional Airport	Abilene
4	ADK	Adak Airport	Adak
5	GUM	Guam International Airport	Agana
6	BQN	Rafael Hernández Airport	Aguadilla
7	CAK	Akron-Canton Regional Airport	Akron
8	ABY	Southwest Georgia Regional Airport	Albany
9	ALB	Albany International Airport	Albany
10	ABQ	Albuquerque International Sunport	Albuquerque
11	AEX	Alexandria International Airport	Alexandria
12	ABE	Lehigh Valley International Airport	Allentown
13	APN	Alpena County Regional Airport	Alpena
14	AMA	Rick Husband Amarillo International Airport	Amarillo
15	ANC	Ted Stevens Anchorage International Airport	Anchorage
16	ATW	Appleton International Airport	Appleton

+ ≡ airline flight airport cancellation

Amarillo → PK

Verde → Atributo

Rosado → FK

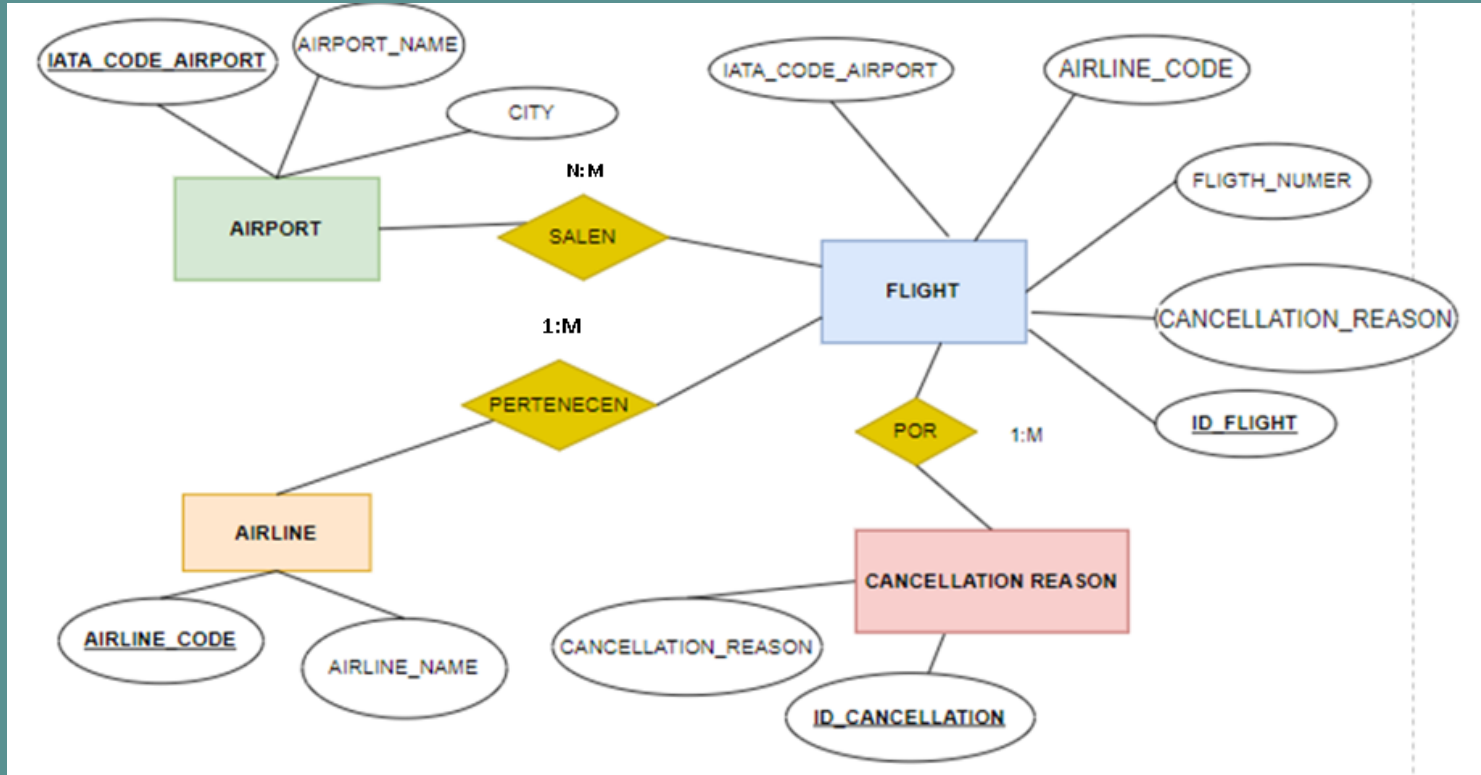
	A	B	C	D	E
1	ID_FLIGHT	AIRLINE_CODE	FLIGHT_NUMBER	IATA_CODE_AIRPORT	CANCELLATION_REASON
2	136ANC	AS	136	ANC	A
3	2459PHX	AA	2459	PHX	B
4	5254MAF	OO	5254	MAF	B
5	2859SGF	MQ	2859	SGF	B
6	5460RDD	OO	5460	RDD	A
7	2926CHS	MQ	2926	CHS	B
8	6457SMX	OO	6457	SMX	A
9	3534ABI	MQ	3534	ABI	B
10	3161XNA	MQ	3161	XNA	B
11	175DCA	AA	175	DCA	B
12	1103LGA	AA	1103	LGA	B
13	1055MCO	AA	1055	MCO	B
14	1180SFO	AA	1180	SFO	B
15	1462ONT	AA	1462	ONT	B
16	865MSO	F9	865	MSO	A

+ ≡ airline **flight** airport cancellation

Hoja de Tabla - “Cancellation” →

A	B
ID_CANCELLATION	CANCELLATION_REASON
A	Condiciones climáticas adversas
B	Problemas técnicos
C	Problemas gremiales

Entrega 2: Diseño del Esquema de la Base de Datos (DER) y Especificaciones

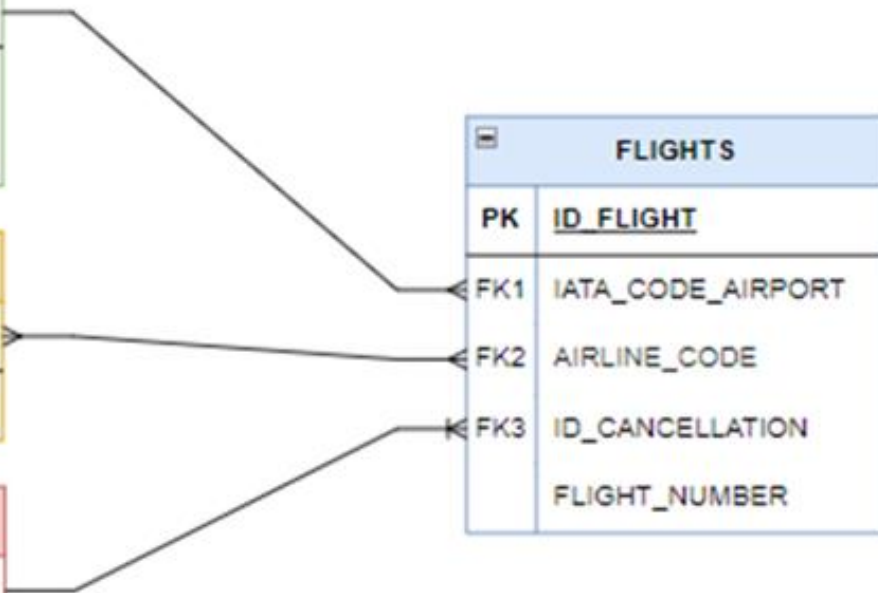


AIRPORT	
PK	<u>IATA_CODE_AIRPORT</u>
	AIRPORT_NAME
	CITY

AIRLINE	
PK	<u>AIRLINE_CODE</u>
	AIRLINE_NAME

CANCELLATION REASON	
PK	<u>ID_CANCELLATION</u>
	CANCELLATION_REASON

FLIGHTS	
PK	<u>ID_FLIGHT</u>
FK1	IATA_CODE_AIRPORT
FK2	AIRLINE_CODE
FK3	ID_CANCELLATION
	FLIGHT_NUMBER



Entidad: Airline

Campo	Tipo	Descripción
<i>AIRLINE_CODE (PK)</i>	<i>varchar(2)</i>	<i>Identificador de aerolínea</i>
<i>AIRLINE_NAME</i>	<i>varchar(50)</i>	<i>Nombre de aerolínea</i>

Entidad: Airport

Campo	Tipo	Descripción
<i>AIRLINE_CODE (PK)</i>	<i>varchar(2)</i>	<i>Identificador de aerolínea</i>
<i>AIRLINE_NAME</i>	<i>varchar(50)</i>	<i>Nombre de aerolínea</i>

Entidad: Flight

Campo	Tipo	Descripción
FLIGHT_NUMBER	<i>varchar(4)</i>	<i>Descripción del número de vuelo</i>
ID_FLIGHT (PK)	<i>varchar(10)</i>	<i>identificador único de vuelo</i>
AIRLINE_CODE (FK)	<i>varchar(2)</i>	<i>Identificador de aerolínea</i>
IATA_CODE (FK)	<i>varchar(3)</i>	<i>Identificador de aeropuerto</i>
CANCELLATION_REASON (FK)	<i>char</i>	<i>Identificador del motivo de cancelación</i>


Entidad: Cancellation_Reason

Campo	Tipo	Descripción
ID_CANCELLATION (PK)	<i>char</i>	<i>Identificador del motivo de cancelación</i>
CANCELLATION_REASON	<i>varchar(50)</i>	<i>Descripción del motivo: A: Condiciones climáticas adversas B: Problemas técnicos C: Problemas gremiales D: Cuestiones de seguridad</i>

Base de Datos en MySQL

```
1 • create database vuelos;
2 • use vuelos;
3 • create table airline(
4     airline_code varchar(2) primary key,
5     airline_name varchar(50)
6 );
7 • create table airport(
8     iata_code varchar(3) primary key,
9     airport_name varchar(80),
10    city varchar(50)
11 );
12 • create table cancellation_reason(
13     id_cancellation char primary key,
14     cancellation_reason varchar(50)
15 );
16 • create table flight(
17     airline_code varchar(2),
18     flight_number varchar(4),
19     iata_code varchar(3),
20     cancellation_reason char,
21     id_flight varchar(10) primary key,
22     foreign key(airline_code) references airline(airline_code),
23     foreign key(iata_code) references airport(iata_code),
24     foreign key(cancellation_reason) references cancellation_reason(id_cancellation)
25 );
```

```
26 • show variables like "secure_file_priv";
27 • load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/airline.csv'
28 into table airline
29 fields terminated by ';'
30 lines terminated by '\n'
31 ignore 1 rows
32 ;
33 • load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/airport.csv'
34 into table airport
35 fields terminated by ';'
36 lines terminated by '\n'
37 ignore 1 rows
38 ;
39 • load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/cancellation_reason.csv'
40 into table cancellation_reason
41 fields terminated by ';'
42 lines terminated by '\n'
43 ignore 1 rows
44 ;
45 • load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/flight.csv'
46 into table flight
```



```
47      fields terminated by ';'
48      lines terminated by '\n'
49      ignore 1 rows
50      ;
51 •    select * from airline;
52 •    select * from airport;
53 •    select * from cancellation_reason;
54 •    select * from flight;
```

[vuelos.sql](#)
(Archivo en drive de MySql)

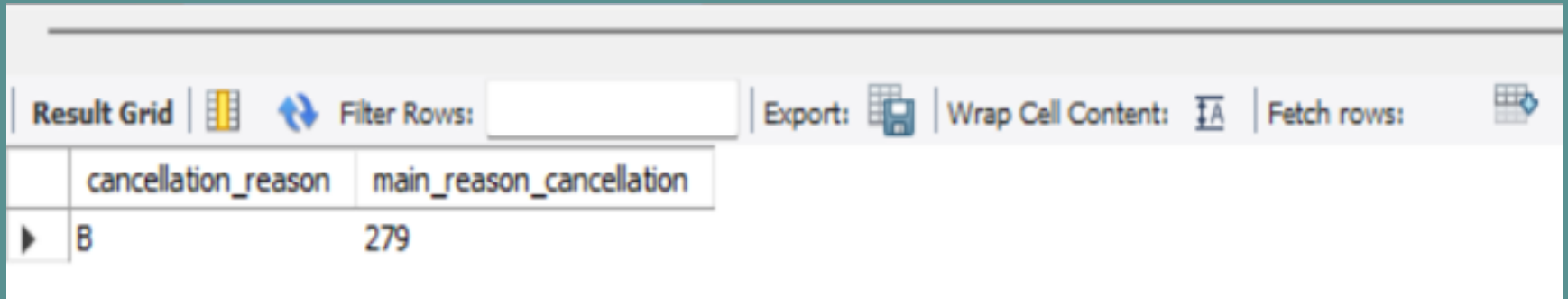
Entrega 3: Análisis Final integrando KPIs y OKRs

A. Análisis de Datos: Consultas SQL Utilizadas:

Consulta 1:

```
• select cancellation_reason, COUNT(*) as number_appearances
  -- selecciona la columna y renombra el contador
from flight
  -- selecciona la tabla vuelos
group by cancellation_reason
  -- agrupa por motivo de cancelacion por lo que contará los motivos repetidos
order by number_appearances desc
  -- ordena los resultados de forma descendiente
limit 1
/* limita el resultado a uno de la consulta
así solo aparecerá el motivo mas frecuente de cancelacion de vuelos
y la cantidad de veces que aparece */
;
```

Resultado 1:



The screenshot shows a web-based data table interface. At the top, there is a toolbar with several icons and labels: 'Result Grid' with a grid icon, a refresh icon, 'Filter Rows:' with an empty input field, 'Export:' with a floppy disk icon, 'Wrap Cell Content:' with a text wrap icon, and 'Fetch rows:' with a grid and arrow icon. Below the toolbar is a table with two columns: 'cancellation_reason' and 'main_reason_cancellation'. The first row of data shows 'B' in the first column and '279' in the second column. A small play button icon is visible to the left of the first cell in the first row.

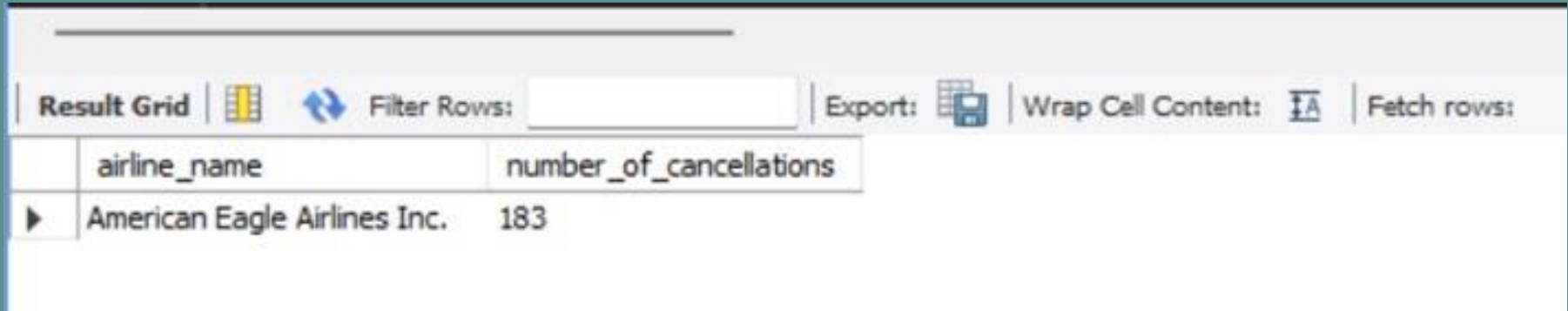
	cancellation_reason	main_reason_cancellation
▶	B	279

De 368 vuelos cancelados en total, 279 corresponden a “B”, es decir, “Problemas Técnicos”.

Consulta 2:

```
81
82 • select a.airline_name, count(*) as number_of_cancellations
83 from flight f
84 -- selecciona el nombre de la aerolínea y cuenta cuántas veces aparece en la tabla de flight
85 join airline a on f.airline_code = a.airline_code
86 -- combina las tablas flight y airline según el airline code
87 where f.cancellation_reason = 'B'
88 group by a.airline_name
89 -- filtra los resultados con motivo de cancelación "b" y los agrupa por nombre de aerolínea
90 order by number_of_cancellations desc
91 limit 1
92 /* ordena de forma descendente y muestra solo el primer resultado
93 obteniendo así el nombre de la aerolínea que más frecuentemente
94 cancela por problemas técnicos */
95 ;
96
```

Resultado 2:



The screenshot shows a web-based data table interface. At the top, there is a toolbar with the following elements: 'Result Grid' with a grid icon, a blue circular refresh icon, a 'Filter Rows:' label followed by an empty text input box, an 'Export:' label with a document icon, a 'Wrap Cell Content:' label with a text icon, and a 'Fetch rows:' label. Below the toolbar is a table with two columns: 'airline_name' and 'number_of_cancellations'. The first row of data shows 'American Eagle Airlines Inc.' and '183'. A small black triangle icon is visible to the left of the first data row.

airline_name	number_of_cancellations
American Eagle Airlines Inc.	183

De 14 aerolíneas, la que más cancelaciones de vuelo tiene por problemas técnicos es la aerolínea de American Eagle Airlines Inc.



OKRs

- **OKR1:** Mejorar la eficiencia del servicio disminuyendo la contratación de aerolíneas con altas tasas de cancelaciones de vuelos.
- **OKR2:** Analizar y comprender las razones detrás de las cancelaciones de vuelos más frecuentes para finales de 2015.

KPIs

- **Frecuencia de cancelaciones por aerolínea:**
Ayuda a las aerolíneas a monitorear y mejorar su desempeño en términos de fiabilidad de sus servicios.
- **Frecuencia de cancelaciones por motivo:**
Se identifica las causas más comunes de interrupciones de vuelos para una aerolínea específica.
- **Tasa de satisfacción del cliente:**
Nivel de satisfacción de los pasajeros con el desempeño de la empresa en cuanto a cancelaciones de vuelos.
- **Costo de las cancelaciones:**
Impacto financiero total de las cancelaciones de vuelos en un día específico. Incluye costos como reembolsos, compensaciones a pasajeros, redireccionamiento de vuelos y pérdida de ingresos.

Conclusiones Finales

Hallazgos del Análisis de la Base de Datos de Vuelos

Mediante el análisis realizado en MySQL Workbench, pudimos identificar que la mayor frecuencia de vuelos cancelados, estaba dada por problemas técnicos en la aerolínea American Eagle Airlines Inc.

En base a éste análisis, la decisión que podría tomar la empresa de viajes es optar por no trabajar con la aerolínea en cuestión.