Aviacion Mexicana: *Facing Post-Pandemic Challenges*

This investigation is brought to you by:

David Ortega Segura

Ruth Falomir de la Pena

Grisell Tavitas Mendoza

Jesus Salas

Are **loyalty programs** a valid strategy for airline companies to overcome **post-pandemic hardships**?

Why is this important?

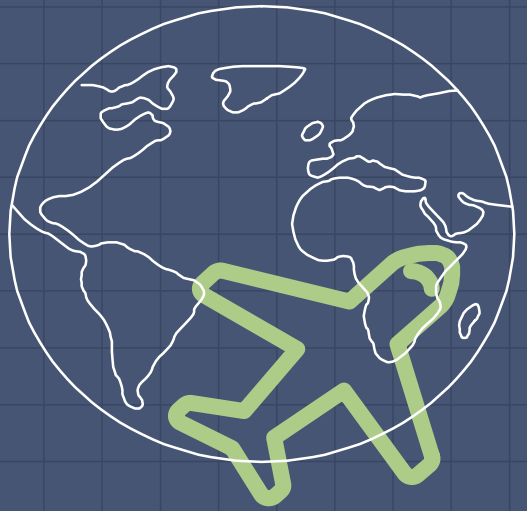
The whole world changed in 2020, but the air transport industry was one of the most affected...

1

Pandemic impact on Air Travel

- According to IATA, 1.8 billion passengers flew in 2020, a **decrease of 60.2%** compared to the 4.5 billion who flew in 2019.
- The total industry passenger **revenues fell by 69%** to $189 billion in 2020, and net losses were $126.4 billion in total.
- On average, there was a **$71.7 loss incurred per passenger** in 2020.

Loyalty Programs



Companies have been actively looking for ways to spend their resources effectively and efficiently. With no dollar to spare, they are looking into many alternatives that have the most impact in recovering their operations.

Are they relevant? How can they help?

80%

of a company's future revenue will come from just 20% of existing clients.

5%

Higher costs to acquire a new customer over keeping an existing one satisfied and happy.

Database

To accomplish the objective of our investigation, we have chose a database that contains the outcomes of 103,904 airline passenger satisfaction survey.

kaggle

Database

The dataset contains the following variables:

sid	id	gender	customer_type	age	type_travel	class_name	flight_distance	departure_delay	arrival_delay	satisfaction
1	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	25	18	neutral or dissatisfied
2	5047	Male	disloyal Customer	25	Business travel	Business	235	1	6	neutral or dissatisfied
3	110028	Female	Loyal Customer	26	Business travel	Business	1142	0	0	satisfied
4	24026	Female	Loyal Customer	25	Business travel	Business	562	11	9	neutral or dissatisfied
5	119299	Male	Loyal Customer	61	Business travel	Business	214	0	0	satisfied

kaggle

Database

Can we predict whether a customer will become loyal?

The analysis could help the airline invest in loyalty programs based on travel rewards

Project Development:
Building a Machine Learning Model

STEP 1 - Decision Making

The model chosen was logistic regression. These models are used to analyze data and, mathematically **determining the probability of new samples belonging to a class.** There are only two possible answers.

STEP 2 - Application

For this investigation, the objective was to **predict whether new purchasing clients will be loyal or disloyal.** The purpose of predicting this outcome is to create better loyalty programs and to dedicate resources to the right customers.

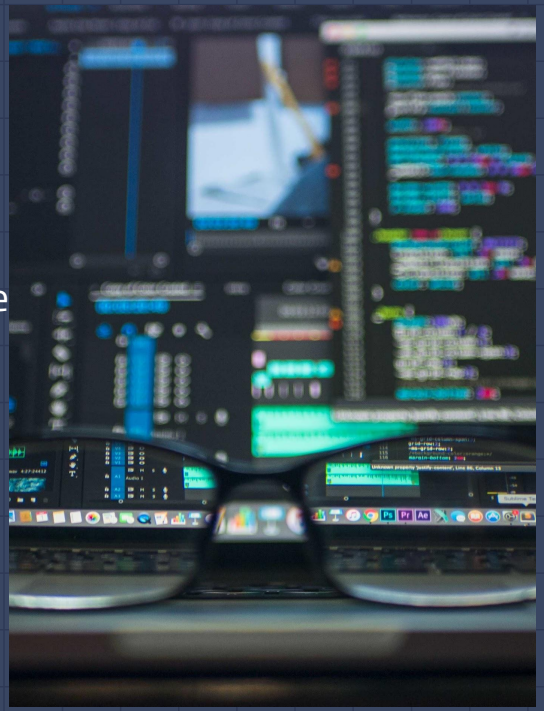
STEP 3 - Validation

After creating the model, it was confirmed that **this particular data could be correctly analyzed using this strategy.**

Data Preprocessing

In order to clean and prepare the data for the model, several steps were established. They go as follows:

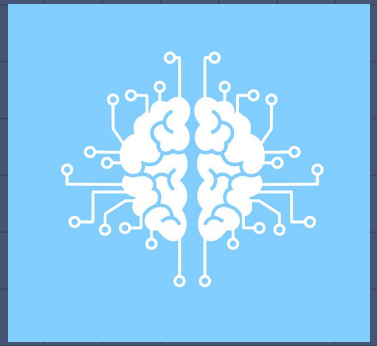
1. Detect missing values with the Pandas DataFrame function isna().
2. Make sure we are using the correct variables data types -Pandas DataFrame dtypes property.
3. Use Scikit-learns LabelEncoder module to transform categorical and text variables into numerical data.
4. Drop the identification id column.
5. Verify the information on the DataFrame, including the index type and columns, non-null values, and memory usage.



Machine Learning

FEATURE ENGINEERING & FEATURE SELECTION

We decided to keep every column except for any that worked as an identifier. Since the rest of the variables are **facts that we consider could have an important impact** on the customer loyalty.



TRAINING & TESTING SETS

We split the dataset into **random** train and test subsets using the Scikit-learn train_test_split module. The training subset will be used in the model to learn from it and the testing subset to assess its performance.

CLASSIFICATION MATRIX

After creating the model, we obtained these results:

	precision	recall	f1-score	support
0	0.92	0.95	0.93	21231
1	0.72	0.63	0.67	4745
accuracy			0.89	25976
macro avg	0.82	0.79	0.80	25976
weighted avg	0.88	0.89	0.89	25976

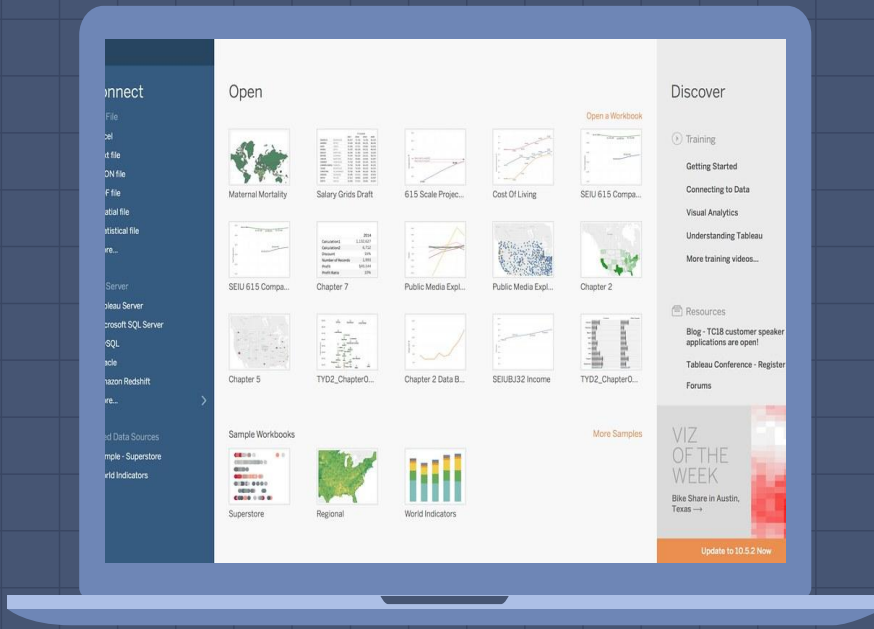
Big picture: How does data look like today?

Why Tableau?

- We choose Tableau to create our dashboard because is a powerful tool with data visualization capabilities and data can be analyzed very quickly.

- One of the most interesting tools that Tableau provides is interactivity. Some of the graphs created can be modified by the viewer using several filters in order to make a deeper analysis.

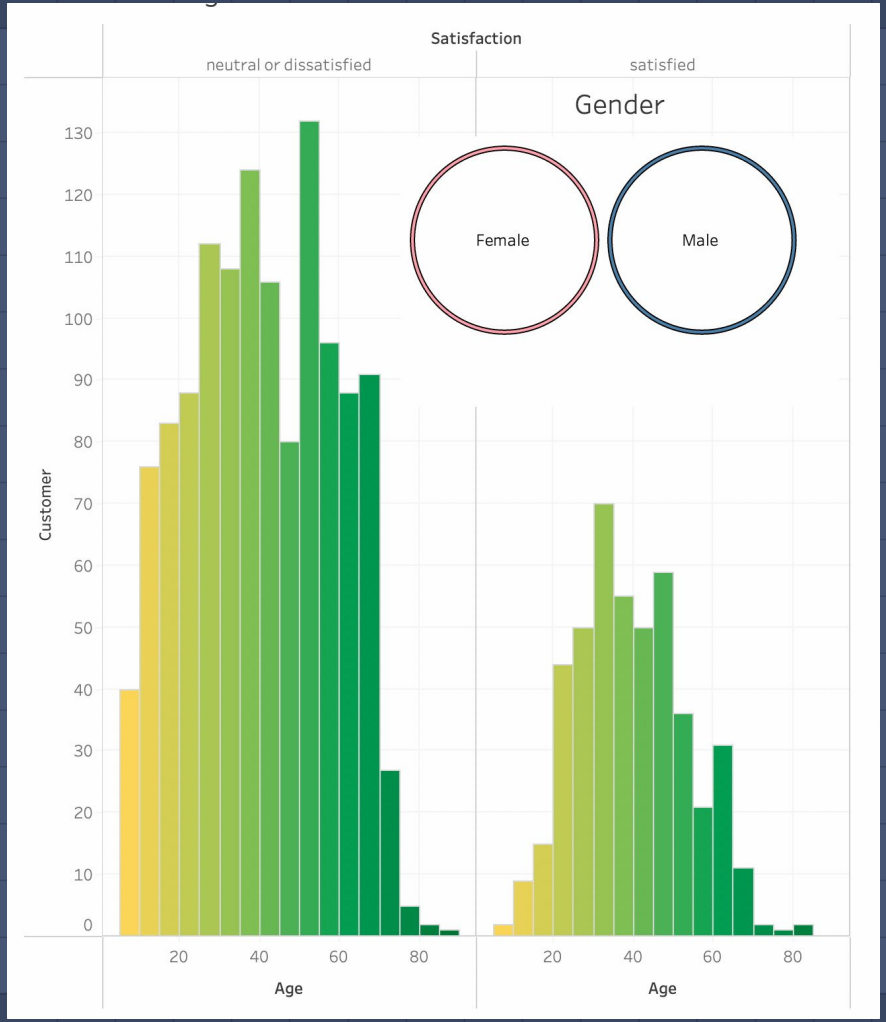
- The tableau dashboard can be viewed publicly here: https://public.tableau.com/app/profile/jesus.salas6470/viz/Dashboard_16464975652980/FinalDashboard



The age distribution between satisfied and dissatisfied compared with the type of class

Passengers between 40- 45 and 20-25 are the largest group and all are Loyal customers.

Males between 30 and 35 years old and females between 50 to 55 years represent the largest group of dissatisfied customers



Comparing the actual data with the machine learning algorithm prediction

Conclusions & Recommendations

As mentioned by my colleagues, the model worked correctly to identify the customer type which will become part of a bigger marketing strategy related to loyalty programs for Mexicana Aviacion.

We recommend as a topic for further research to look into the features that have the 20-25 age group dissatisfied in order to retain them as loyal clients in the future.

THANKS!