



# SC1015

## Airline Reviews

Group 6 FCEA: Timothy, Param, Wong Yi Jie

# Table of contents



01

Introduction

02

Problem Statement

03

EDA

04

Machine Learning

05

Insights

06

Our Recommendations

# Background

By the end of 2024,

- Airline industry net profits are expected to reach **\$25.7 billion**
- Overall traveler numbers will reach **4.0 billion**



With such high passenger numbers and market share to be obtained, airlines are undoubtedly pushing to increase their prominence in the market

# Background

Airlines aim to entice passengers by providing a high quality of service throughout the passenger journey

Some of the common factors include:

1. Seat Comfort,
2. Cabin Staff Service,
3. Food & Beverages,
4. Ground Service,
5. Inflight Entertainment,
6. Wifi & Connectivity



Hence, we were curious as to what factors would normally affect a passenger's overall rating and whether they recommend an airline or not

# Chosen Dataset

The dataset we chose is from

<https://www.kaggle.com/datasets/juhibhojani/airline-reviews>



Airline Name	Overall Rating	Review Title	Review Date	Verified	Review	Type Of Traveller	Seat Type	Route	Date Flown	Seat Comfort	Cabin Staff Service	Food & Beverages	Ground Service	Inflight Entertainment	Wifi & Connectivity	Value For Money	Recommended
Silver Airways	1	"I never made it to the plane"	1st August 2020	TRUE	"I never made it to the plane"	Solo Leisure	Economy Class	San Juan to St. I.	June 2020							2.0	no
Silver Airways	1	"worst customer service ever"	3rd June 2020	TRUE	Absolutely the	Family Leisure	Economy Class	San Juan to St. I.	June 2020							1.0	no
Silver Airways	1	"never fly this airline ever again"	18th April 2020	TRUE	My friends and	Solo Leisure	Economy Class	San Juan to St. I.	April 2020							1.0	no
Singapore Air	9	"Great service and food"	22nd July 2023	FALSE	Thank you to	Couple Leisure	Business Class	Singapore to Singapore	July 2023	5.0	5.0	5.0	5.0	5.0	4.0	5.0	yes
Singapore Air	1	"by far the worst customer"	19th July 2023	FALSE	Singapore Air	Couple Leisure	Economy Class	Davao to New Y.	January 2023	1.0	1.0	1.0	1.0	1.0	1.0	1.0	no
Singapore Air	9	"Service was excellent as"	17th July 2023	TRUE	SG Silver Kris	Family Leisure	Business Class	Singapore to New	April 2023	5.0	5.0	4.0	5.0	4.0	5.0	5.0	yes
Singapore Air	6	"not as good I was expecti"	16th July 2023	FALSE	I was looking to	Solo Leisure	Premium Econo	New York to Frai	May 2023	4.0	4.0	4.0	3.0	4.0	4.0	5.0	yes
Singapore Air	2	"ask my wife to do it by hei"	9th July 2023	TRUE	We had this un	Family Leisure	Economy Class	Dempasar to Kar	July 2023	4.0	4.0	4.0	1.0	4.0	4.0	4.0	no
Singapore Air	1	"I was pleasantly surprised"	4th July 2023	TRUE	My first time o	Solo Leisure	Economy Class	Auckland to Sing	July 2023	4.0	5.0	5.0	5.0	5.0	5.0	4.0	yes
Singapore Air	8	"In-flight service was excel"	28th June 2023	TRUE	We did not rec	Family Leisure	Business Class	Hanoi to Singa	April 2023	5.0	5.0	4.0	3.0	4.0	4.0	4.0	yes
Singapore Air	9	"really nice for a sub 2-hou"	27th June 2023	TRUE	The Silver Kris	Family Leisure	Business Class	Singapore to Ho	April 2023	5.0	5.0	4.0	5.0	4.0	4.0	5.0	yes
Singapore Air	7	"In-flight service was quite"	27th June 2023	TRUE	I was supposed	Family Leisure	Business Class	New York to Sing	April 2023	5.0	5.0	4.0	2.0	4.0	4.0	5.0	yes
Singapore Air	2	"needed a GF Vegan meal"	23rd June 2023	FALSE	When booking	Solo Leisure	Economy Class	Brisbane to Lon	May 2023	3.0	5.0	1.0	1.0	3.0	3.0	2.0	no
Singapore Air	6	"not professional behavior"	20th June 2023	TRUE	I find it equali	Couple Leisure	Business Class	Singapore to Mil	June 2023	2.0	3.0	3.0	3.0	2.0	4.0	3.0	no
Singapore Air	5	"Worst business class seat"	16th June 2023	TRUE	Worst business	Couple Leisure	Business Class	Manchester to S	June 2023	1.0	4.0	4.0	5.0	5.0	5.0	3.0	no
Singapore Air	7	"They still have high stand"	16th June 2023	FALSE	I flew to Singa	Couple Leisure	Economy Class	London Heathro	May 2023	4.0	5.0	3.0	4.0	4.0	4.0	3.0	yes
Singapore Air	6	"seats were hard and unco"	12th June 2023	FALSE	We travelled re	Couple Leisure	Business Class	London to Auckl	February 2023	1.0	5.0	4.0	4.0	4.0	4.0	3.0	no
Singapore Air	9	"special praise for Air Stew"	9th June 2023	TRUE	I would like to	Family Leisure	Premium Econo	Singapore to Los	June 2023	5.0	5.0	3.0	5.0	4.0	5.0	5.0	yes
Singapore Air	2	"another 6 weeks for the re"	7th June 2023	FALSE	Not such a gre	Solo Leisure	Economy Class	Singapore to Bai	April 2023				2.0			2.0	no
Singapore Air	6	"rather disappointed"	1st June 2023	TRUE	The First flight	Business	First Class	Singapore to Syl	May 2023	3.0	4.0	3.0	4.0	3.0	4.0	4.0	yes
Singapore Air	9	"a very nice experience"	30th May 2023	TRUE	I flew with SG	Solo Leisure	Business Class	Auckland to Sing	May 2023	3.0	5.0	4.0	5.0	5.0	5.0	4.0	yes
Singapore Air	7	"provided the very best"	30th May 2023	TRUE	From start to fi	Solo Leisure	First Class	Singapore to Los	May 2023	3.0	5.0	5.0	5.0	5.0	5.0	4.0	yes
Singapore Air	1	"The suite very comfortable"	30th May 2023	TRUE	Suites on the A	Solo Leisure	First Class	London to Singa	May 2023	5.0	5.0	5.0	4.0	5.0	5.0	5.0	yes
Singapore Air	5	"totally disappointing"	29th May 2023	TRUE	The A380 layou	Business	Business Class	New York JFK to	April 2023	1.0	3.0	1.0	1.0	2.0	2.0	2.0	yes
Singapore Air	1	"my luggage was damaged"	28th May 2023	TRUE	Very nice and	Business	Economy Class	Johannesburg to	May 2023	3.0	3.0	2.0	1.0	4.0	4.0	2.0	no
Singapore Air	1	"Compensation should hav"	26th May 2023	TRUE	Late 2.5 hours	Couple Leisure	Premium Econo	Manchester to S	May 2023	1.0	1.0	1.0	1.0	2.0	2.0	1.0	no
Singapore Air	2	"Very poor service"	26th May 2023	TRUE	Very poor servi	Business	Business Class	Los Angeles to S	May 2023	3.0	1.0	1.0	4.0	4.0	4.0	1.0	no
Singapore Air	8	"Service and crew were exc"	22nd May 2023	TRUE	Always loved S	Solo Leisure	Business Class	Frankfurt to Sing	February 2023	3.0	5.0	4.0	5.0	4.0	4.0	4.0	yes
Singapore Air	5	"Cabin staff didn't seem ve"	22nd May 2023	TRUE	HRD-SRD was	Couple Leisure	Economy Class	Tokyo to London	May 2023	2.0	3.0	1.0	2.0	2.0	1.0	2.0	no
Singapore Air	4	"not recommend this airline"	22nd May 2023	TRUE	We used this ai	Family Leisure	Economy Class	Kochi to Singa	May 2023	3.0	2.0	2.0	2.0	2.0	1.0	3.0	no
Singapore Air	4	"wide ranging cost cutting"	20th May 2023	TRUE	What has happ	Couple Leisure	Business Class	London to Singa	May 2023	2.0	5.0	2.0	3.0	4.0	3.0	2.0	no
Singapore Air	4	"the service and staff was	14th May 2023	TRUE	It's been a lon	Solo Leisure	Economy Class	Brisbane to Sing	April 2023	4.0	2.0	2.0	3.0	5.0	5.0	3.0	no
Singapore Air	1	"helpful, courteous and pol	12th May 2023	TRUE	SG28 SRH-LAN	Couple Leisure	Economy Class	Singapore to Los	May 2023	5.0	5.0	5.0	4.0	5.0	5.0	5.0	yes
Singapore Air	1	"another great flight with S"	12th May 2023	TRUE	SG287 OPS-OP	Couple Leisure	Economy Class	Dempasar to Sing	May 2023	4.0	5.0	4.0	5.0	5.0	5.0	5.0	yes
Singapore Air	1	"cabin crew truly went abo	11th May 2023	TRUE	Despite some	Couple Leisure	Business Class	Istanbul to Sing	May 2023	5.0	5.0	5.0	5.0	5.0	5.0	5.0	yes
Singapore Air	2	"50 minutes for the priority"	8th May 2023	TRUE	This was a ver	Solo Leisure	Economy Class	Singapore to Del	April 2023	3.0	5.0	1.0	2.0	1.0	1.0	1.0	no
Singapore Air	4	"late by up to 90 mins"	4th May 2023	TRUE	Flight I had w	Solo Leisure	Business Class	Frankfurt to New	April 2023	2.0	4.0	2.0	2.0	3.0	3.0	1.0	no
Singapore Air	8	"revitalise the quality of inf	2nd May 2023	TRUE	22 April flight	Couple Leisure	Economy Class	Singapore to Sei	April 2023	3.0	5.0	3.0	4.0	4.0	4.0	4.0	yes
Singapore Air	1	"a really great crew"	28th April 2023	TRUE	Used Singapore	Solo Leisure	Economy Class	Dempasar to Bar	April 2023	5.0	5.0	3.0	5.0	4.0	4.0	4.0	yes
Singapore Air	3	"refused to offer me any cc	28th April 2023	TRUE	My flight on 14	Business	Economy Class	San Francisco to	March 2023	4.0	5.0	4.0	1.0	5.0	3.0	3.0	no
Singapore Air	9	"This airline is worth the mi	23rd April 2023	FALSE	This airline is	Couple Leisure	Business Class	Brisbane to Joh	March 2023	4.0	4.0	4.0	4.0	4.0	4.0	4.0	yes

# Problem Statement

**Analyse airline reviews to identify factors affecting passenger ratings and recommendations.**

- Highlight changes in consumer preferences from pre to post-COVID.
- Evaluate factors like '**Seat Comfort**', '**Cabin Staff Service**', '**Food & Beverages**', '**Ground Service**', '**Inflight Entertainment**', '**Wifi & Connectivity**', against '**Overall\_Rating**' and '**Recommended**'
- Provide **actionable insights for airlines** to improve their ratings/recommendation.



# 01

# Data Preparation



# Data Cleaning

## Dropped Columns (~~Red~~ strikethrough)

- We have decided to columns “Unnamed”, “Aircraft” and “Route”.

## Reasons

- The unnamed column is just a index column for the rest of the data (**Irrelevant**)
- The Route column just tells us the starting point and destination. (**Irrelevant**)
- Aircraft just tells us the Airplane Model (**Irrelevant**)

RangeIndex: 23171 entries, 0 to 23170

Data columns (total 20 columns):

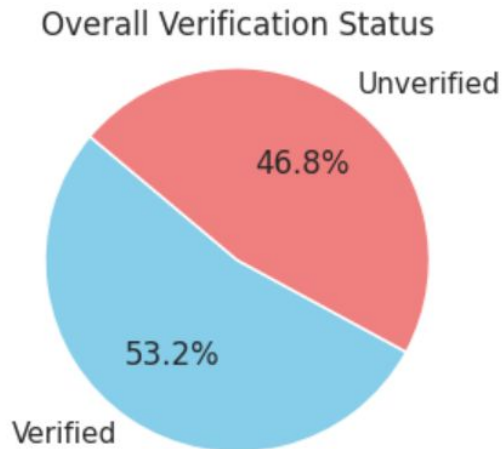
#	Column	Non-Null Count	Dtype
0	<del>Unnamed: 0</del>	23171 non-null	int64
1	Airline Name	23171 non-null	object
2	Overall_Rating	23171 non-null	object
3	Review_Title	23171 non-null	object
4	Review Date	23171 non-null	object
5	Verified	23171 non-null	bool
6	Review	23171 non-null	object
7	<del>Aircraft</del>	7129 non-null	object
8	Type Of Traveller	19433 non-null	object
9	Seat Type	22075 non-null	object
10	<del>Route</del>	19343 non-null	object
11	Date Flown	19417 non-null	object
12	Seat Comfort	19016 non-null	float64
13	Cabin Staff Service	18911 non-null	float64
14	Food & Beverages	14500 non-null	float64
15	Ground Service	18378 non-null	float64
16	Inflight Entertainment	10829 non-null	float64
17	Wifi & Connectivity	5920 non-null	float64
18	Value For Money	22105 non-null	float64
19	Recommended	23171 non-null	object



# Data Cleaning

## Dropping Unverified Reviews

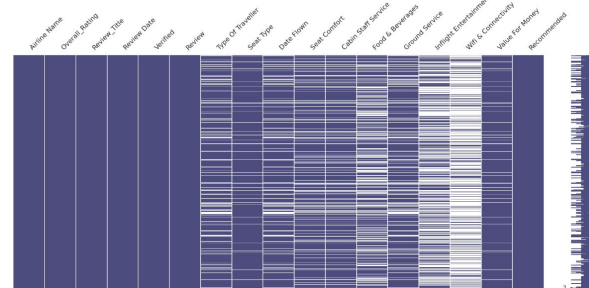
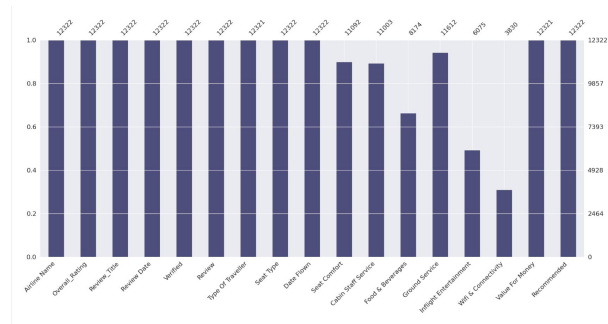
We have also decided to drop those data are **unverified** because we want to **reduce the risk** of incorporating misleading or fraudulent feedback, providing a reliable basis for our conclusions.



# Data Cleaning

## External Library : Missingno for Visualisation

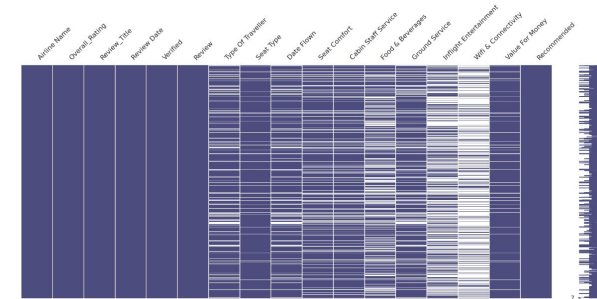
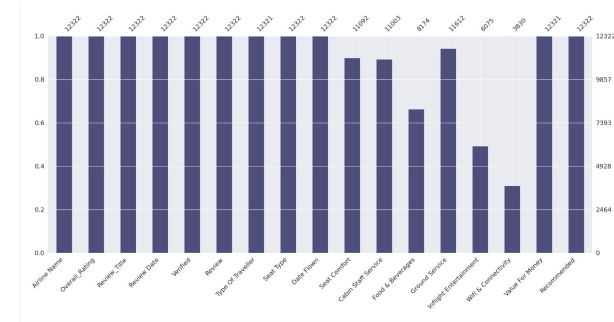
- Missingno provides a small toolset of flexible and easy-to-use **missing data visualizations** and utilities that allows you to get a **quick visual summary** of the completeness (or lack thereof) of your dataset
- **White Space** represents the NaN data within the Dataset



# Data Cleaning

## Summary Analysis of NaN values from Missingno

- Percentage of NaN values in Seat Comfort: **9.98%**
- Percentage of NaN values in Cabin Staff Service: **10.70%**
- Percentage of NaN values in Food & Beverages: **33.66%**
- Percentage of NaN values in Ground Service: **5.76%**
- Percentage of NaN values in Inflight Entertainment: **50.70%**
- Percentage of NaN values in Wifi & Connectivity: **68.92%**



# Data Cleaning

## Comparison of the Two Dataframes we created

- The **verifiedairlineReview\_filled** data frame consists of **all reviews** regardless of whether they did not indicate a rating for any of the service options (**LHS**) with all the **null values filled with 0**.
  - This was done to retain airlines that are Low Cost Carriers and even full service airlines that do not operate with the full suite of services
- The **verifiedairlineReview\_dropped** only has reviews in which all the criterias had a **non NA rating (RHS)**
  - This was the main dataset used in the ML analysis to prevent any unwanted **skew** in our results

# 02

## Exploratory Data Analysis



# Comparing Average Rating of Most Reviewed Airlines

## Top 3 Airlines in **verifiedairlineReview\_filled**

- China Southern Airlines
- Hainan Airlines
- Qatar Airlines

## Top 3 Airlines in **verifiedairlineReview\_dropped**

- Qatar Airways
- Emirates
- Jet Airways

This shows among the full service airlines, Qatar, Emirates and the now defunct Jet Airways emerge as the most rated airlines. However, a high total review count does not equal to a high average rating.

## **verifiedairlineReview\_filled**

	Total_Reviews	Average_Rating
Airline Name		
China Southern Airlines	97	6.309278
Hainan Airlines	97	4.298969
Garuda Indonesia	93	3.870968
Qatar Airways	92	3.978261
Aeroflot Russian Airlines	90	2.811111
...	...	...
Somon Air	1	1.000000
Rhein-Neckar Air	1	1.000000
Lubeck Air	1	1.000000
Libyan Airlines	1	4.000000
Qazaq Air	1	9.000000

## **verifiedairlineReview\_dropped**

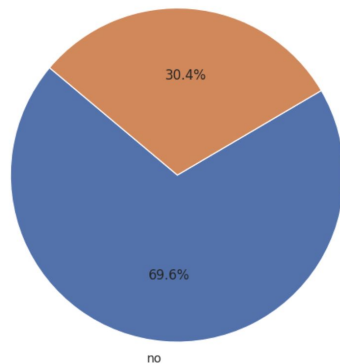
	Total_Reviews	Average_Rating
Airline Name		
Qatar Airways	51	4.039216
Emirates	42	2.833333
Jet Airways	41	1.951220
El Al Israel Airlines	40	2.925000
Singapore Airlines	39	4.000000
...	...	...
Azur Air Germany	1	1.000000
Solomon Airlines	1	1.000000
StarFlyer	1	1.000000
TAR Aerolineas	1	1.000000
Africa World Airlines	1	1.000000

# Recommendation Analysis

In **both** datasets, a great majority of the reviews are probably **unhappy** passengers who would **not recommend** their airline and among full service carriers the **likelihood** of being recommended further **decreases**

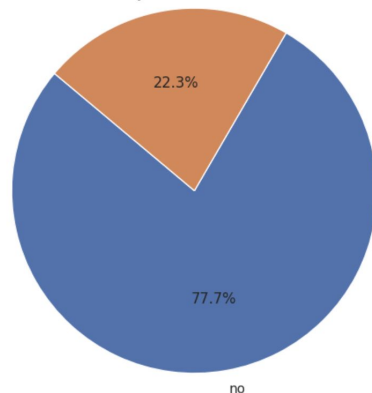
## verifiedairlineReview\_filled

Distribution of Recommended Status



## verifiedairlineReview\_dropped

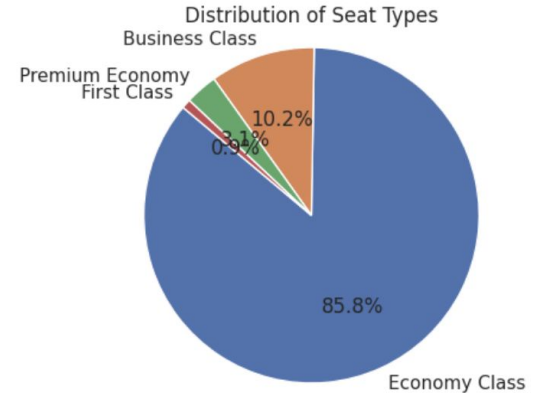
Distribution of Recommended Status



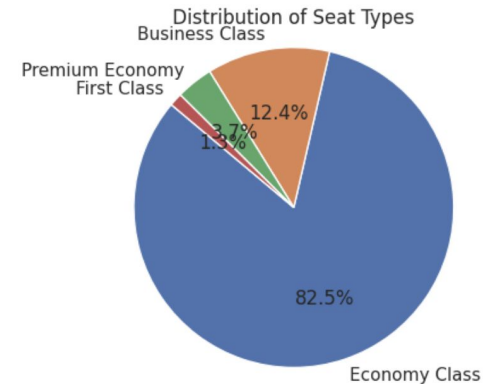
# Seat Type Analysis

The percentage of people travelling in the four classes of travel are rather similar across both datasets but since full service airlines often have more seats in the premium cabin, there is a **slight increase in reviews** of the premium cabins as shown by the decrease in economy class reviews (**85.8% vs 82.5%**)

## verifiedairlineReview\_filled



## verifiedairlineReview\_dropped



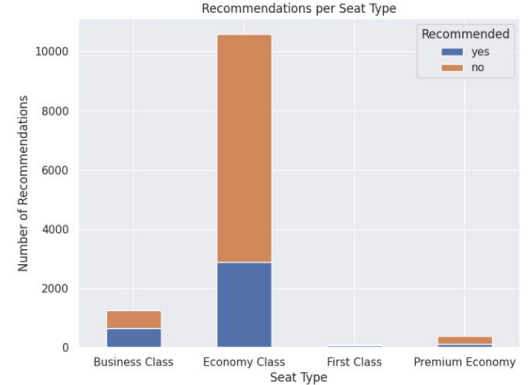


# Recommendation by Seat Type

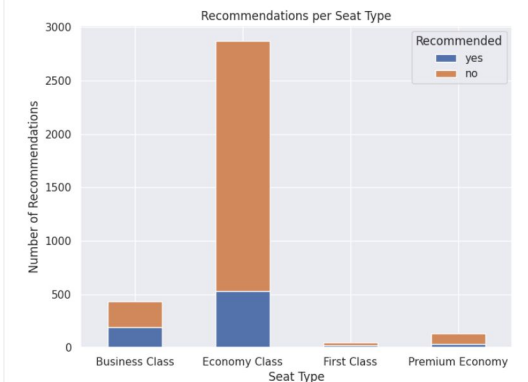
## verifiedairlineReview\_filled

- Economy class passengers are **least likely** to recommend an airline. (% Yes of Total - **27%**)
- First class there is an **increasing trend in likelihood** that passengers would recommend an airline. (% Yes of Total - **60%**)

## verifiedairlineReview\_filled



## verifiedairlineReview\_dropped



## verifiedairlineReview\_dropped

- Economy class passengers are still **least likely** to recommend an airline. (% Yes of Total - **18%**)
- First class there is an **increasing trend in likelihood** that passengers would recommend an airline. (% Yes of Total - **52%**)

# Mean Overall Rating by Seat Type

## verifiedairlineReview\_filled

- Business Class - **4.107**
- First Class - **3.363**

## verifiedairlineReview\_dropped

- Business Class - **3.407**
- First Class - **3.391**

People on **business class** tend to rate their experience **higher or equal than** those in **first class** which is interesting because normally one would expect first class to have an ultra premium service that people would be greatly satisfied with.

## verifiedairlineReview\_filled

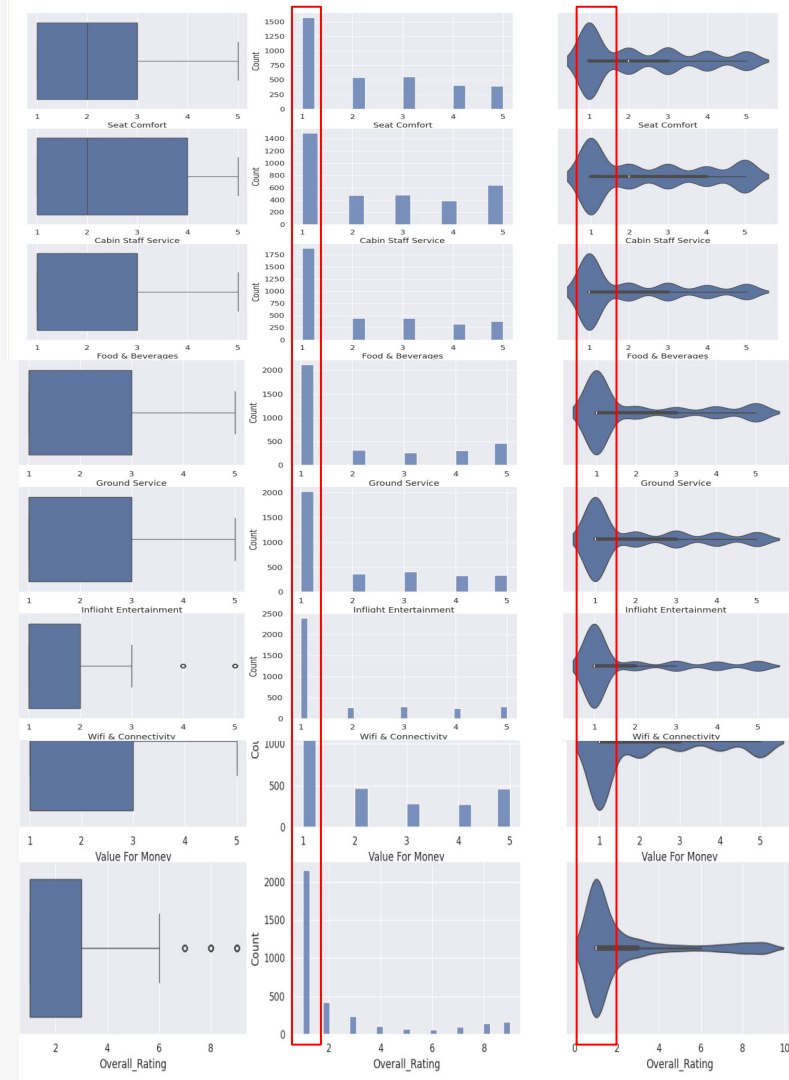


## verifiedairlineReview\_dropped



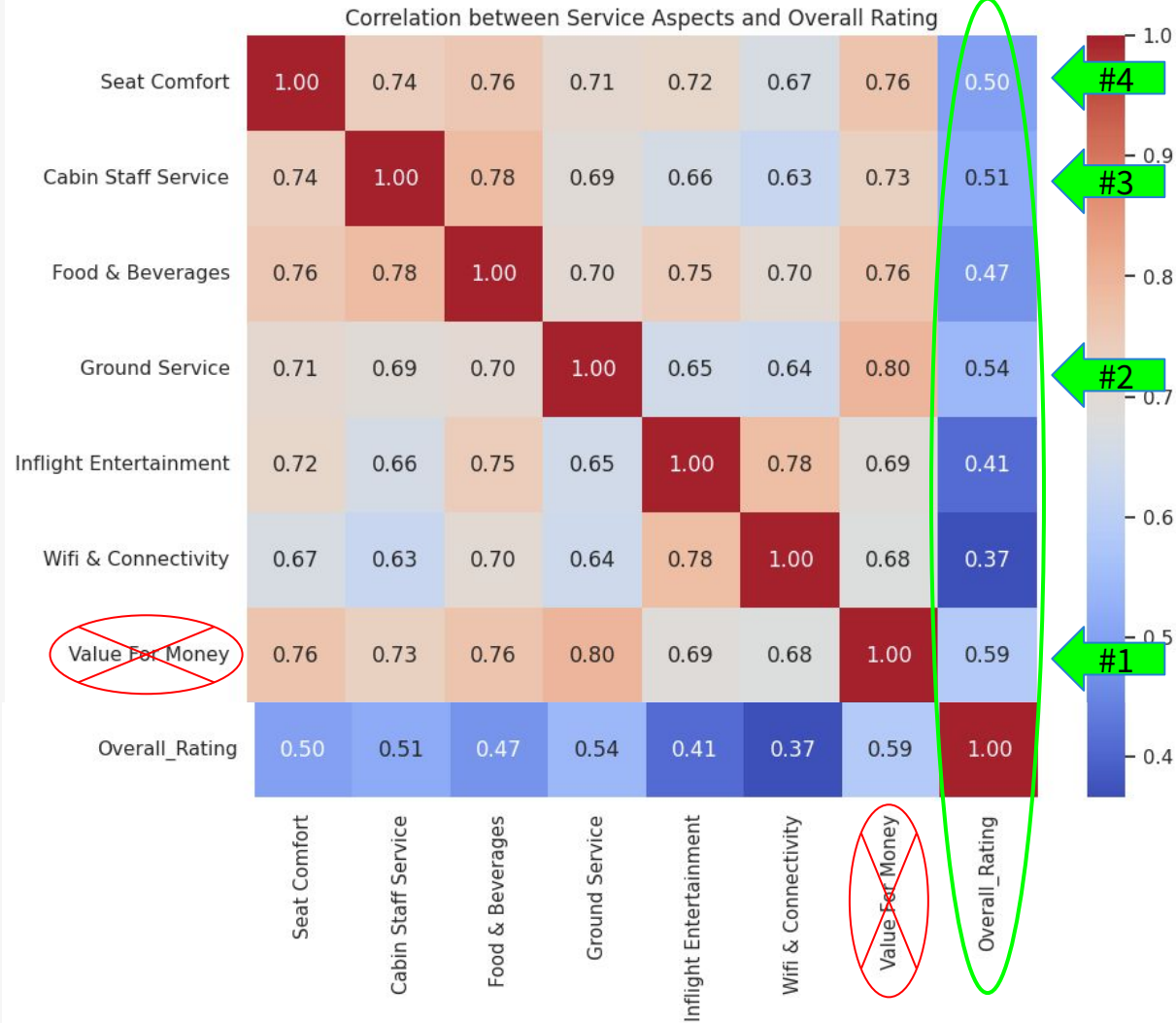
# Visualization of all **Service Aspect Ratings** and **Overall Rating**

- Disproportionate amount of '1' ratings given
  - Likely due to human nature
  - More likely to leave bad reviews instead of good reviews



# Correlation matrix of all Variables that were rated and Overall Rating

- **Green arrows** represent variables that have the **highest correlation** with **Overall Rating**
- **Value for Money** **dropped**
  - As it is a highly subjective variable



# 03

## Machine Learning Techniques



# Decision Tree

- A flowchart-like tree structure where each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label.
- Used to identify which features among Seat Comfort, Cabin Staff Service, Food & Beverages, Ground Service, Inflight Entertainment, and Wifi Connectivity most influence the Overall Rating of airlines and whether an airline is recommended by a passenger.
- **Max\_depth=3**, was used when evaluating against overall rating and **max\_depth = 4** was used when evaluating against recommendation emphasizing readability and interpretability of graph

# Random Forest

- Chose Random Forest as our second model as it is an ensemble of decision trees, and it typically improves predictive accuracy and controls over-fitting.
- Used these parameters when setting our model (`n_estimators=300`, `max_depth=10`, `min_samples_split=5`, `min_samples_leaf=1`, `max_features='auto'`, `random_state=42`)
- As shown by our classification test accuracy, when Recommended was the target variable the test data accuracy improved in our random forest model compared to our decision tree model from **0.9367816091954023 to 0.9741379310344828** and when Overall Rating was the target variable the classification accuracy improved from **0.6057471264367816 to 0.7863984674329502**.

# XGBoost Regression & Classification (as a predictor model)

- Provide airline companies with models that would allow them to see predicted **Overall Rating** and **Recommendation state** when given different combination of service aspect ratings
  - Allow airline companies to see **which aspect to prioritize** to give the best possible **Overall Rating** and **Recommendation**
- XGBoost instead of SKlearn as it can handle complex relationships better with a higher predictive performance

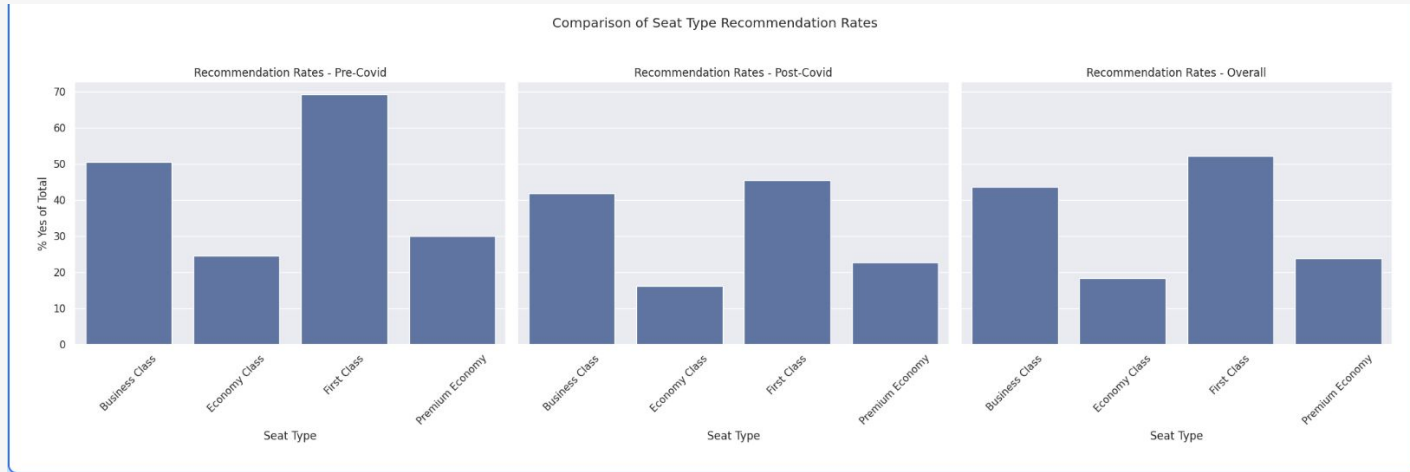


# 04

## Data Driven Insights



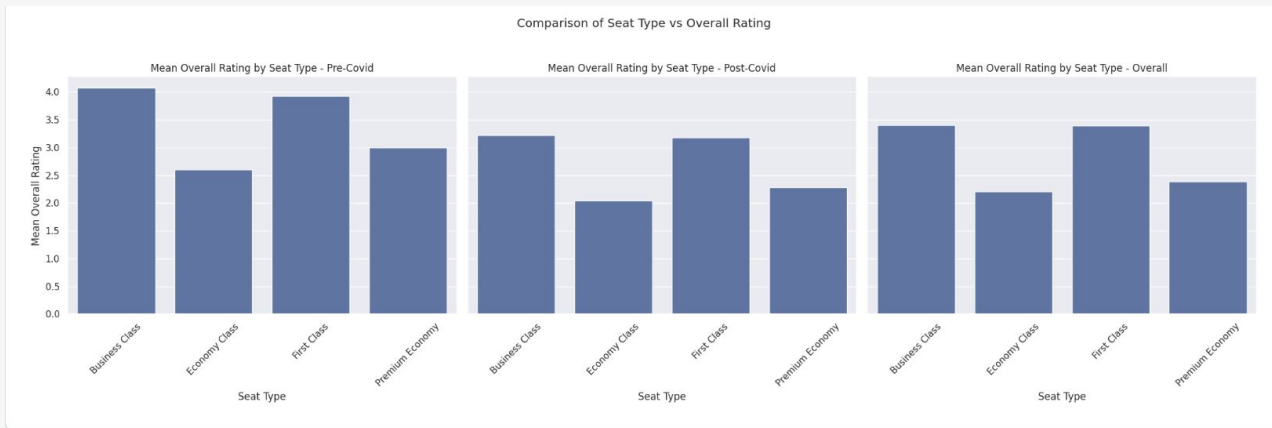
# Recommendation based on Seat Type



**Economy Class:** Has lowest recommendation rates overall, but there is a slight decrease in the post-Covid period which could be due to impact of service changes due to the pandemic.

**Business Class & First Class:** It seems to have consistently high recommendation rates, but the post-Covid drop could be attributed to the changes in service due to health protocols as well which would have fallen short of people's expectations.

# Overall Rating based on Seat Type



**Economy Class:** Has lowest mean overall rating throughout, but there is a slight decrease in the post-Covid period which could be due to impact of service changes due to the pandemic.

**Business Class & First Class:** It seems to have consistently high mean overall rates, but the post-Covid drop could be attributed to the changes in service due to health protocols as well which would have fallen short of people's expectations.

# Services Importance to Overall Rating

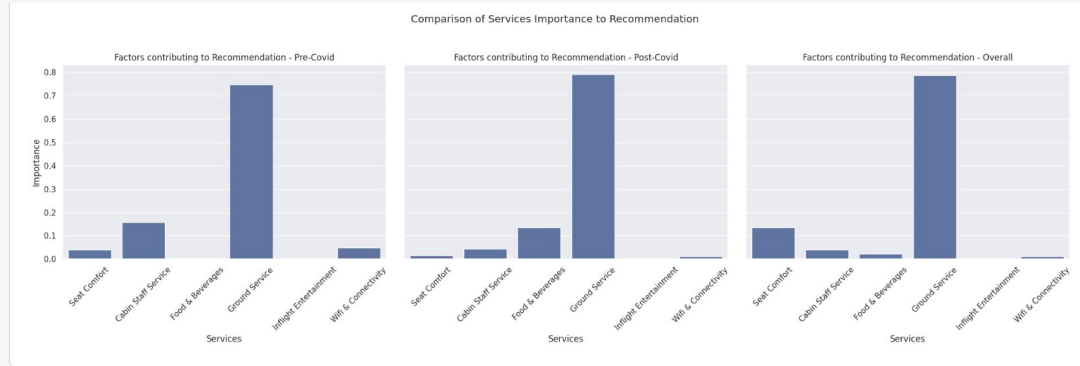


**Ground service** stands out as the **most important factor** for **overall rating** throughout. While in the overall dataset, seat comfort and wifi stand out as runner up factors but the pre and post covid dataset shows that cabin staff service might be important as well

In order to boost overall rating, airlines should consider:

1. Maintain the high importance of ground services. Make ground services easily available and help adequately with customers' needs
2. Investing in comprehensive training programs to ensure staff can effectively handle passenger requests and provide a enjoyable and comfortable journey for the passengers
3. While entertainment is less critical, comfort and connectivity are increasingly important. Enhancing seat comfort, perhaps with more spacious seating arrangements or better cushioning, and improving wifi services could be key areas of improvement.

# Services Importance to Recommendation



**Ground service** once again stands out as the **most important factor** for **recommendation** throughout. While in the overall dataset, seat comfort and cabin staff service stand out as runner up factors but the post covid dataset shows that food and beverages might be important too

In order to boost recommendation numbers, airlines should consider:

1. Enhancing customer experience of by focusing on ground services.
2. Improving in-flight dining experiences which has now potentially got renewed interest
3. Improve or promote their cabin staff service as it was an important factor pre covid but now its relatively low so boosting this might be a differentiating factor for airlines.
4. Maintaining or enhancing seat comfort could improve passenger experience, especially as travel norms and expectations continue to evolve post-pandemic.

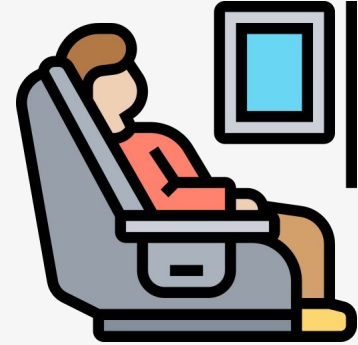
# Our Final Judgement

## Seat Types that yield the best reviews for Airlines:

**Recommendation** -> 1) First Class, 2) Business Class, 3) Premium Economy, 4) Economy

**Overall Rating** -> 1) Business Class, 2) First Class, 3) Premium Economy, 4) Economy

\* However, should airlines want to improve their recommendation and overall rating numbers they should focus on improving their economy class product as that is the segment where passengers have the most complaints and most passengers in.



## Top Services to Focus on for Airlines (Our Final Judgement)

**Recommendation** -> 1) Ground Service, 2) Seat Comfort, 3) Food & Beverages

**Overall Rating** -> 1) Ground Service, 2) Cabin Staff Service, 3) Seat Comfort

\* This does not just follow the overall importance. It accounts for any increase in importance post COVID suggesting that it might be for important for airlines to start focusing on these factors as well.



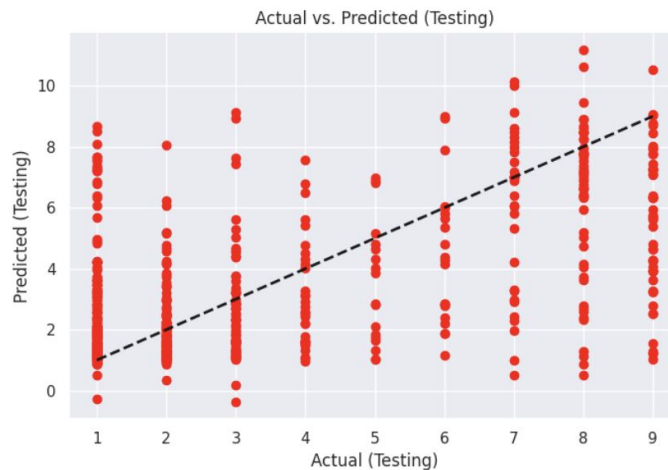
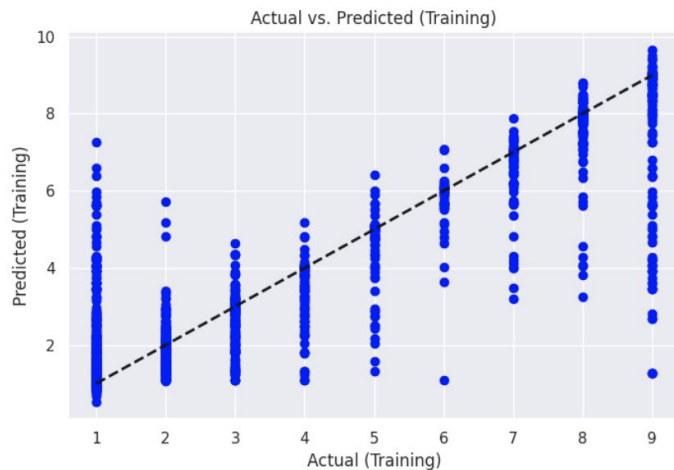


# SC1015

## Airline Reviews

Thank you!

# XGBoost Regression



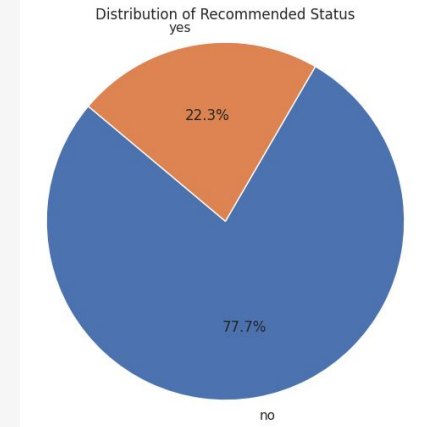
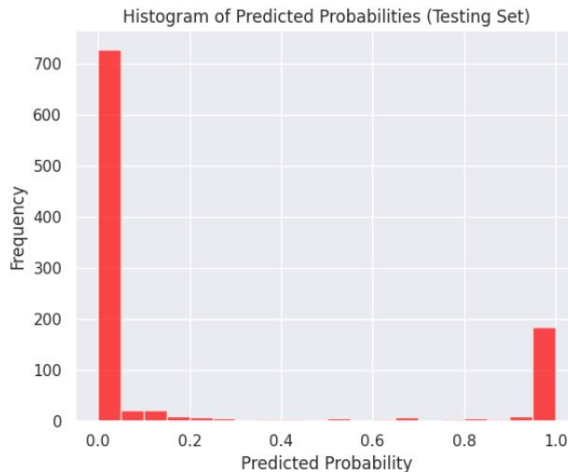
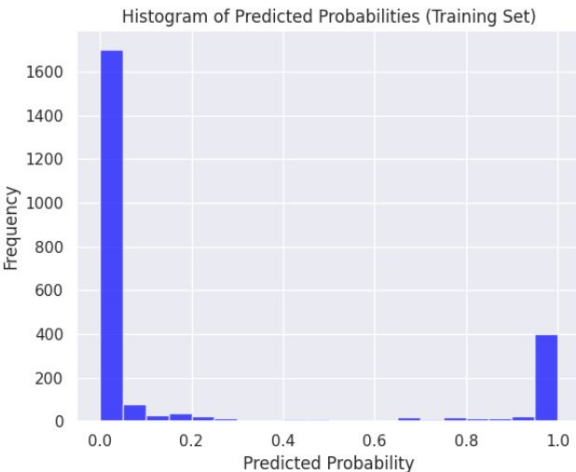
Train Mean Squared Error: 1.1353331823235342  
Test Mean Squared Error: 3.4170394220668703

Entered Seat Comfort rating: 5.0  
Entered Cabin Staff Service rating: 3.0  
Entered Food & Beverages rating: 3.0  
Entered Ground Service rating: 5.0  
Entered Inflight Entertainment rating: 1.0  
Entered Wifi & Connectivity rating: 1.0  
**Predicted Overall Rating: 7.9146285**

Enter Seat Comfort rating (1-5):



# XGBoost Classification



Train Accuracy: 0.9843814221126181

Test Accuracy: 0.9406130268199234

Entered Seat Comfort rating: 5.0

Entered Cabin Staff Service rating: 3.0

Entered Food & Beverages rating: 3.0

Entered Ground Service rating: 5.0

Entered Inflight Entertainment rating: 1.0

Entered Wifi & Connectivity rating: 1.0

Predicted Recommendation: Yes

Enter Seat Comfort rating (1-5):