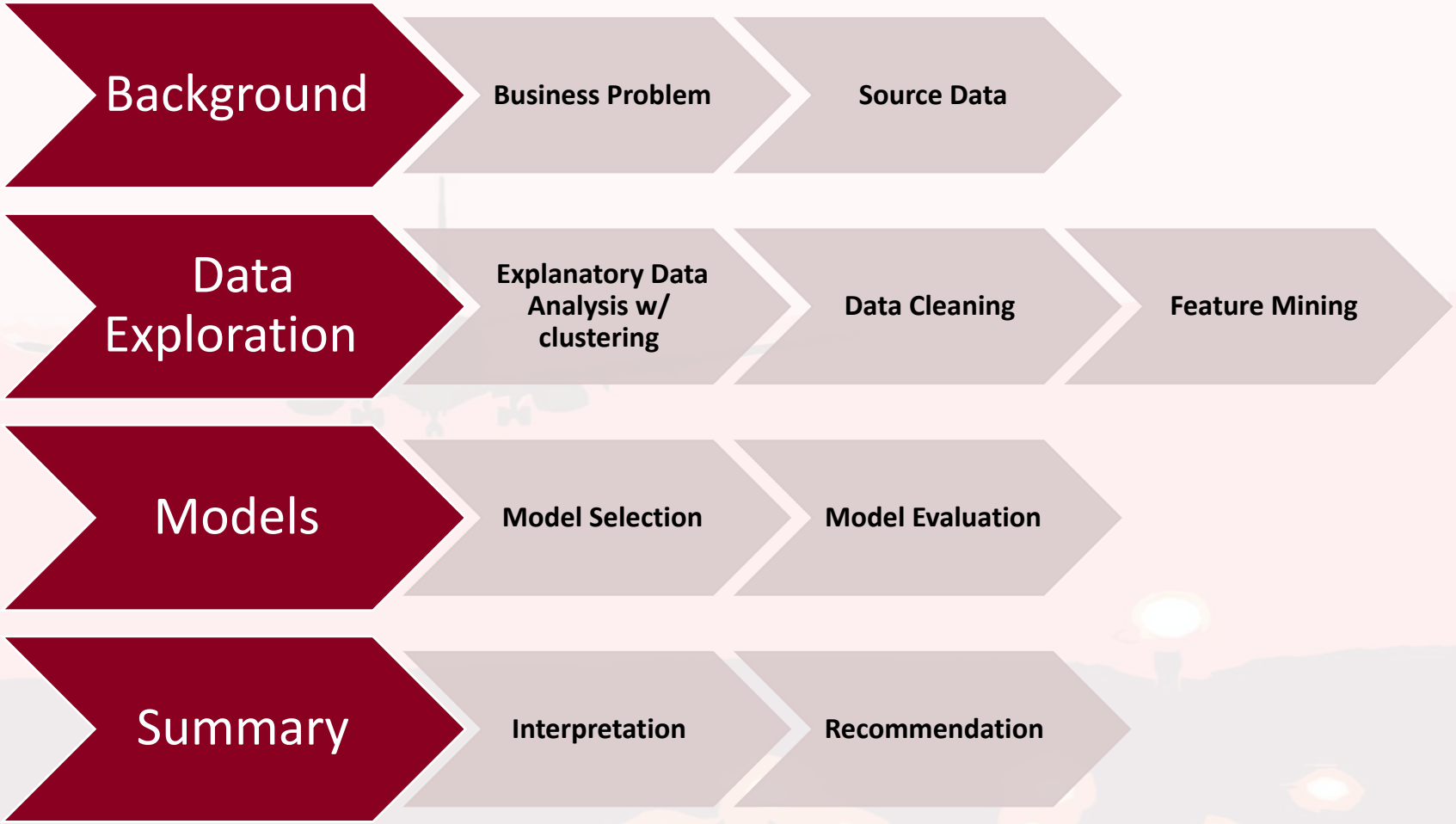
Airline Passenger Satisfaction

What factors lead to customer satisfaction for an Airline?

Guangze Wang Kaicheng Zhang
Meng Yang Wanxuan Zhang Zhiyi Zhao
*In alphabetical order

Agenda


Background
Business Problem
Source Data
Data Exploration
Explanatory Data Analysis w/ clustering
Data Cleaning
Feature Mining
Models
Model Selection
Model Evaluation
Summary
Interpretation
Recommendation

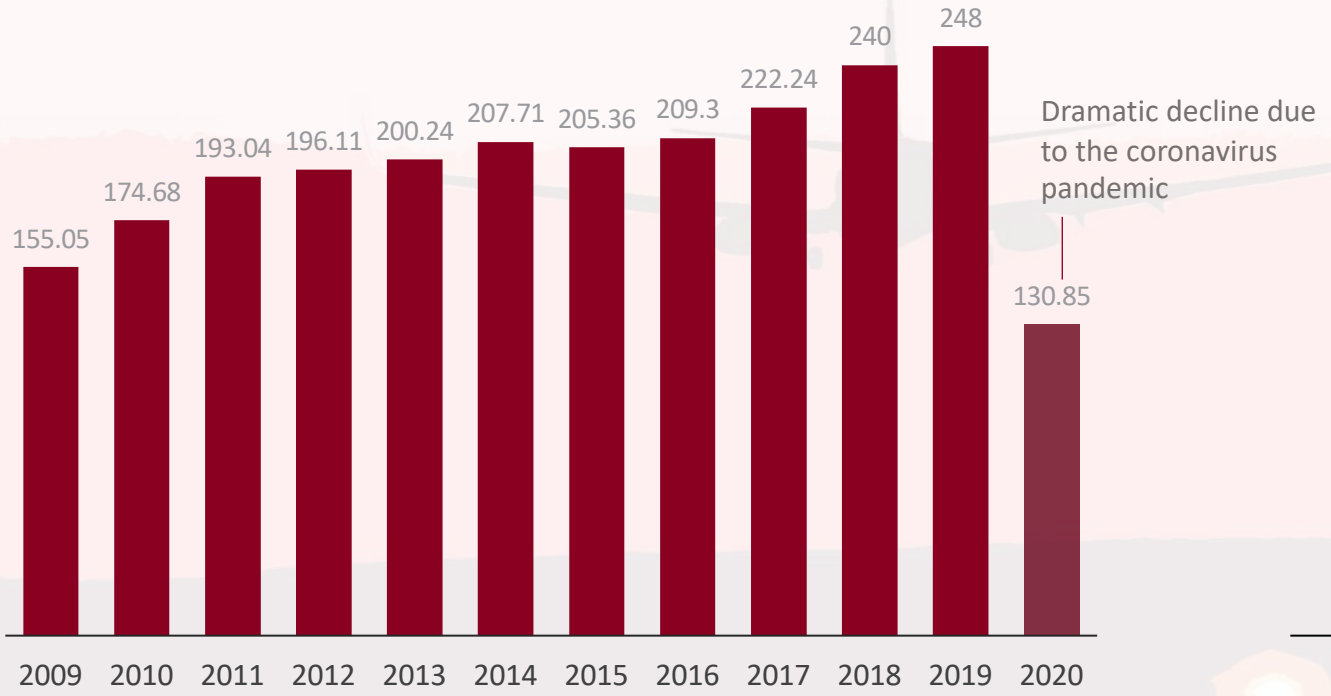
Understand the Business Problem

Customer satisfaction is always top of mind for airlines

THE UNIVERSITY OF CHICAGO

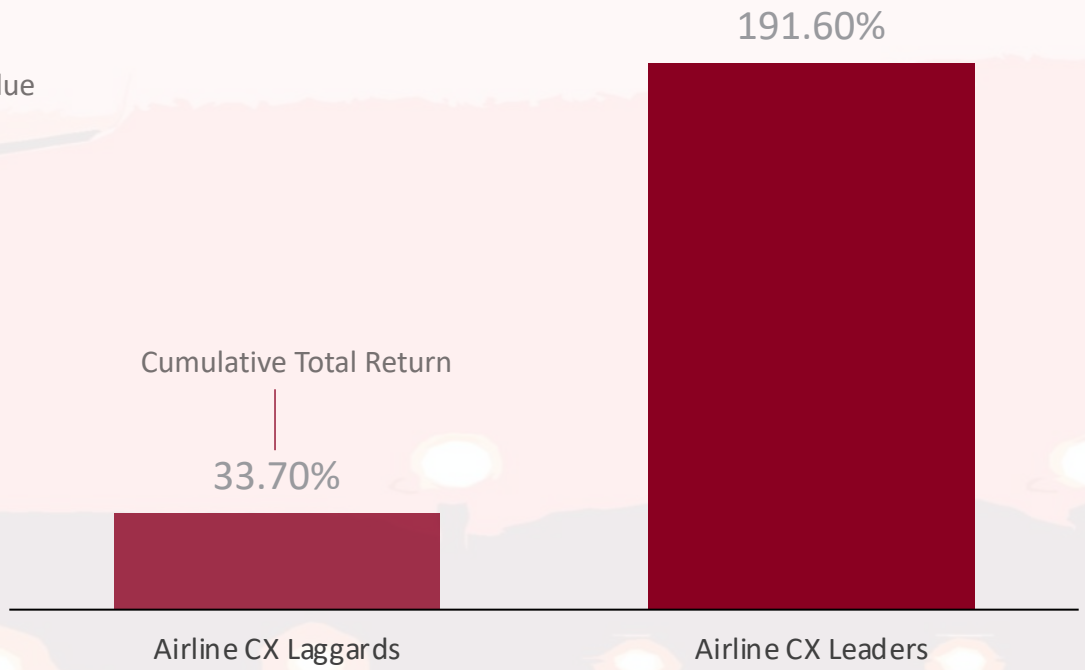
Total operating revenue streams of U.S. airlines from 2009 to 2020 (in billion U.S. dollars)

U.S. airlines urge to recover from the global pandemic



10-Year Stock Performance of Customer Experience(CX) Leaders vs. Laggards (2011-2020)

Airline Customer Experience Leaders Outperform



* Source: statista watermark

Understand the Business Problem

Value Of Our Business Case: the big picture

THE UNIVERSITY OF CHICAGO

Detect unsatisfied flight with high accuracy + Recommend service upgrade + Improve both sales and market share = Increase Economic Value

Increase Passenger Satisfaction

Build Customer Loyalty

Achieve Business Success

Source data

A relatively balanced dataset with mostly hierarchical categorical features

THE UNIVERSITY OF CHICAGO

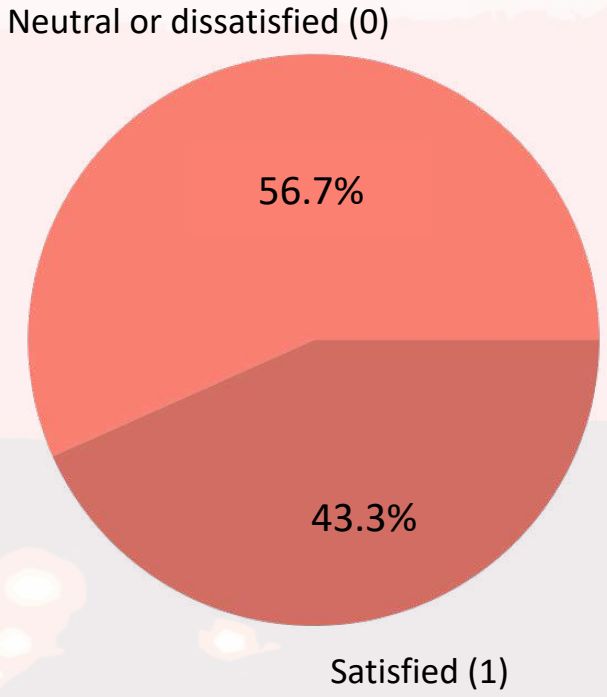
id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/ Arrival time convenient	Ease of Online booking	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction
70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	3	...	5	4	3	4	4	5	5	25	18.0	neutral or dissatisfied
110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	2	...	5	4	3	4	4	4	5	0	0.0	satisfied

103904 instances
28 features

Train size: 0.8
Test size: 0.2

Target variable: satisfaction

numerical features: 5
categorical features: 22



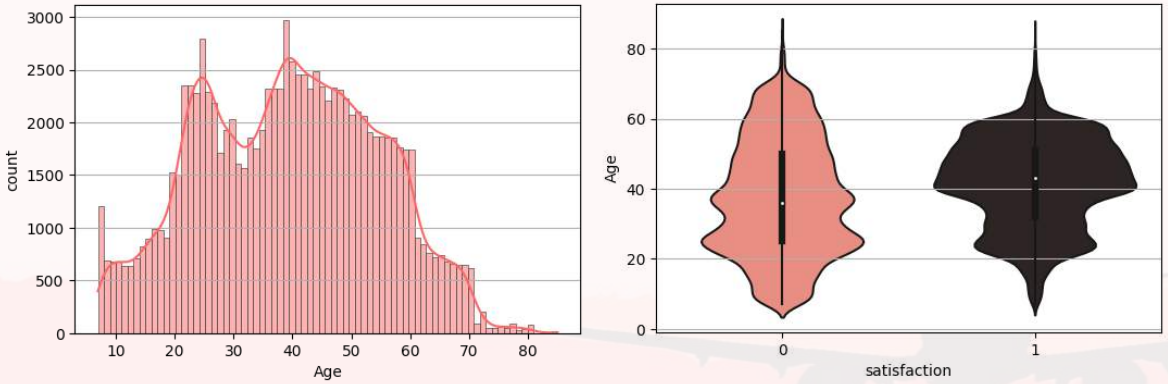
Exploratory data analysis

Numeric features: passenger demographic & flight info

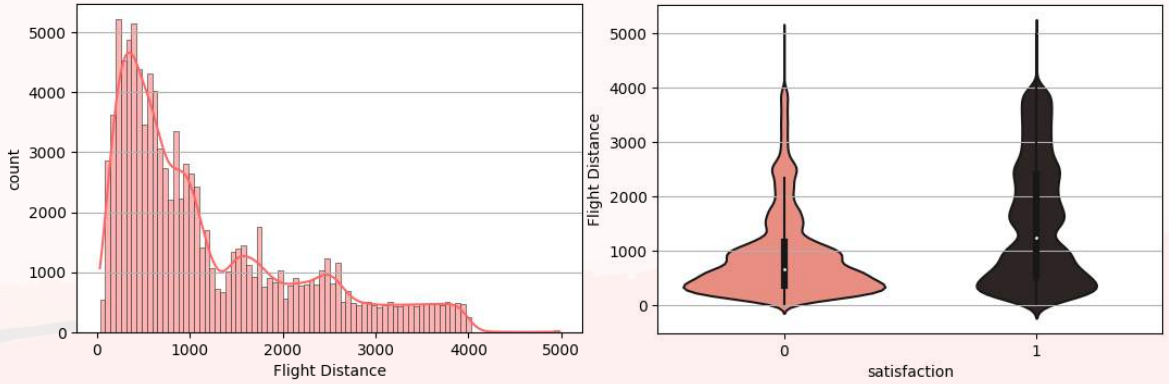
THE UNIVERSITY OF CHICAGO

Findings: Middle-aged passengers & Long-distance flight passengers tend to satisfy more.

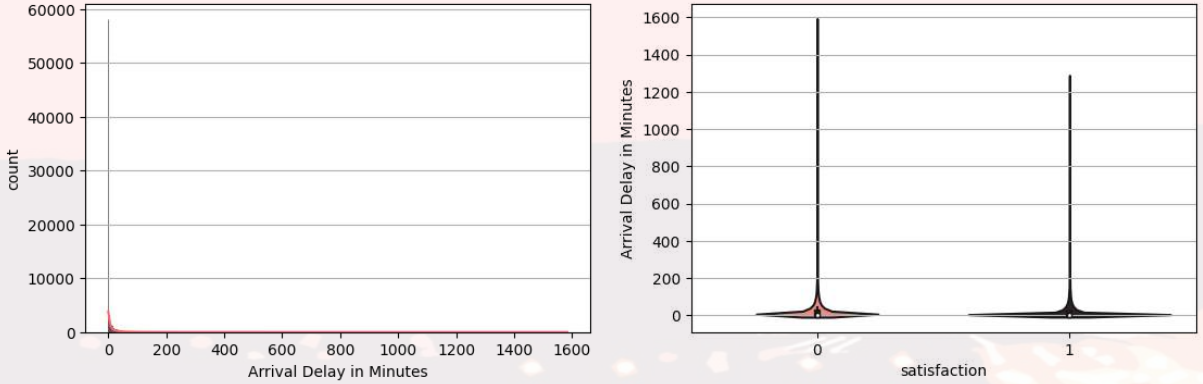
Age



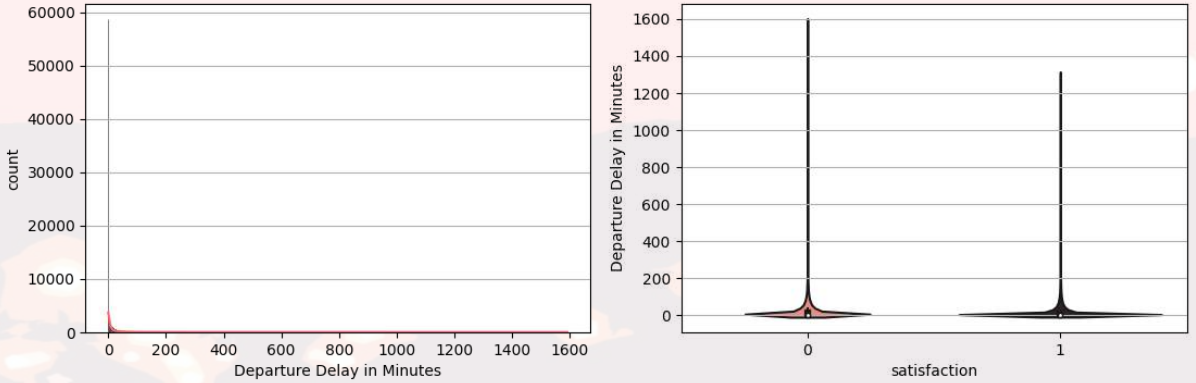
Flight Distance



Arrival Delay in Minutes



Departure Delay in Minutes

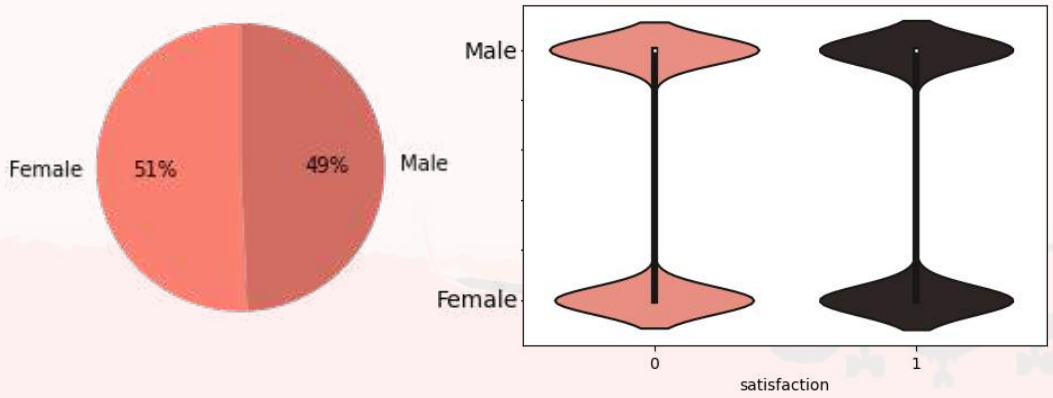


Exploratory data analysis

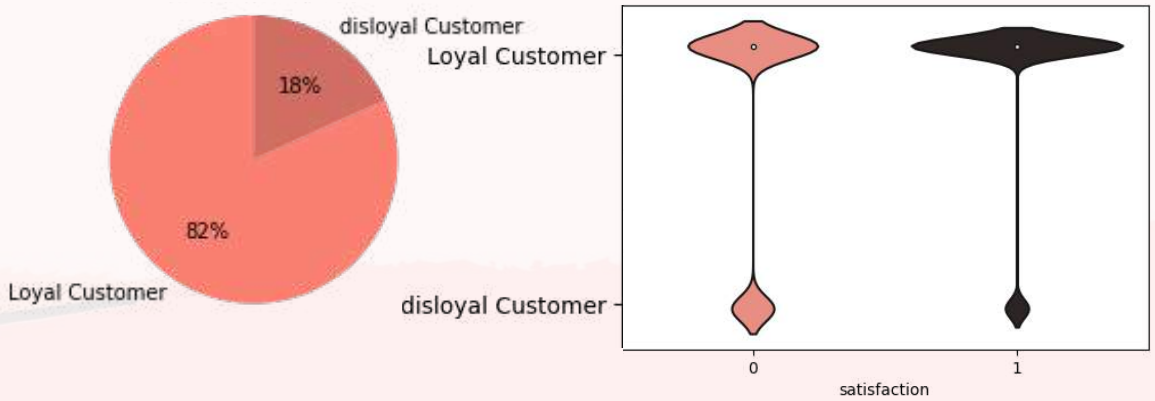
Categorical features: passenger demographic & flight info

Findings: Loyal customers, business travelers & Business class passengers tend to satisfy more.

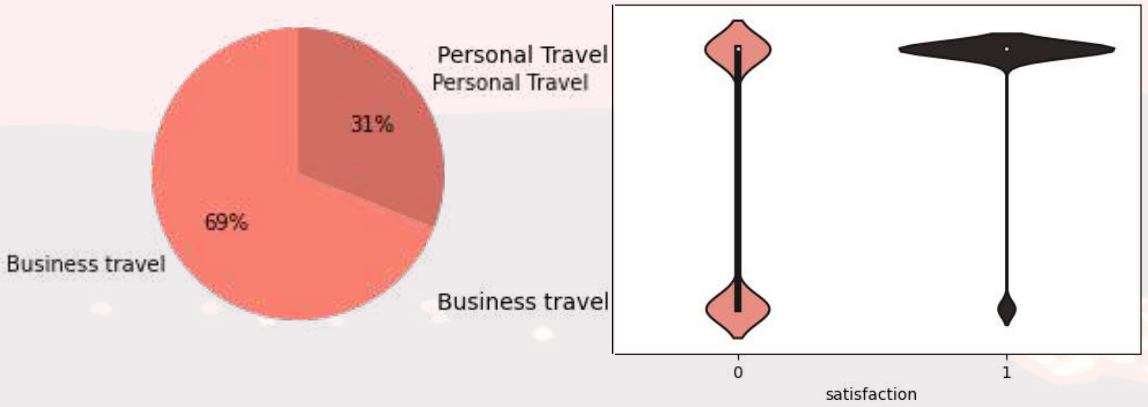
Gender



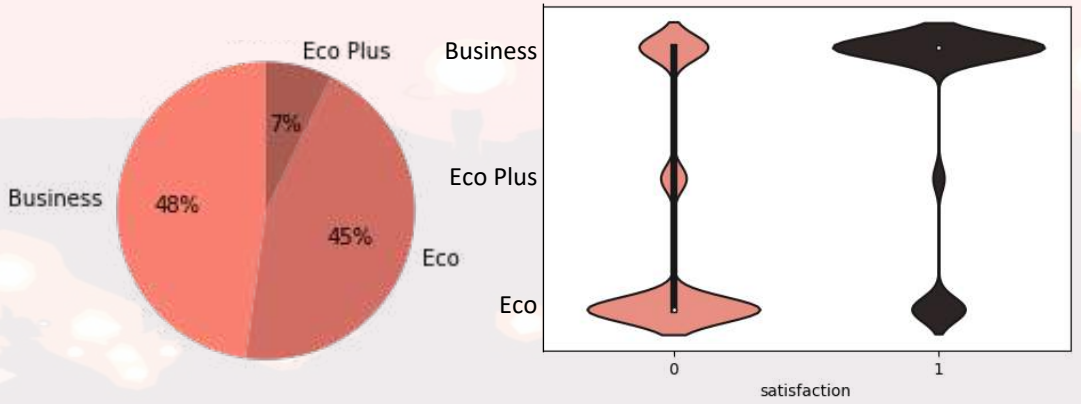
Customer Type



Type of Travel



Class

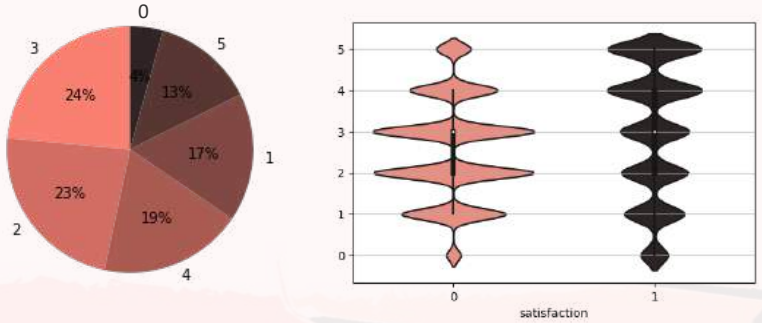


Exploratory data analysis

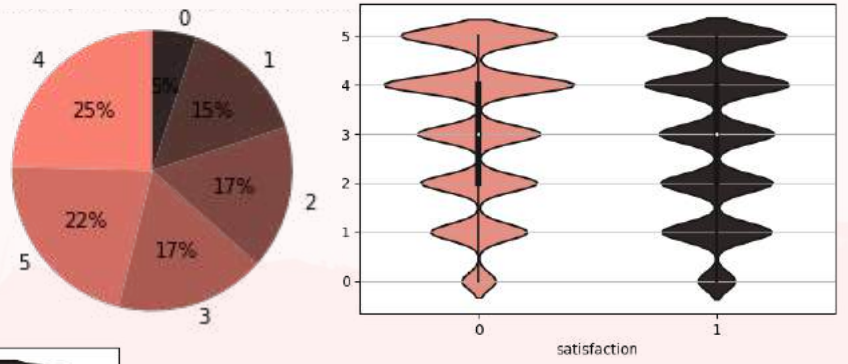
Hierarchical categorical features: before getting into the plane

Findings: Easier online booking and boarding satisfy passengers more.

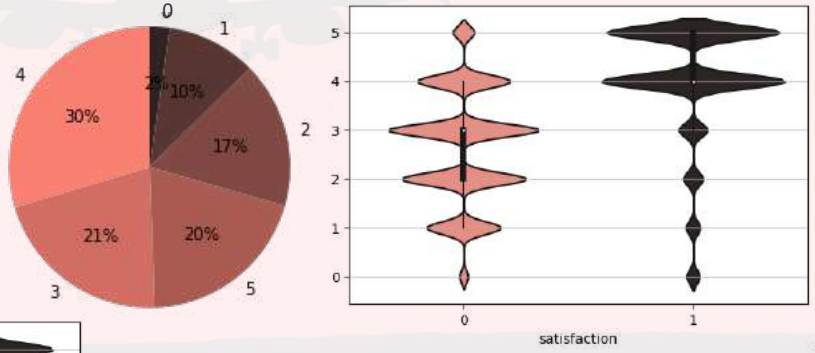
Ease of Online booking



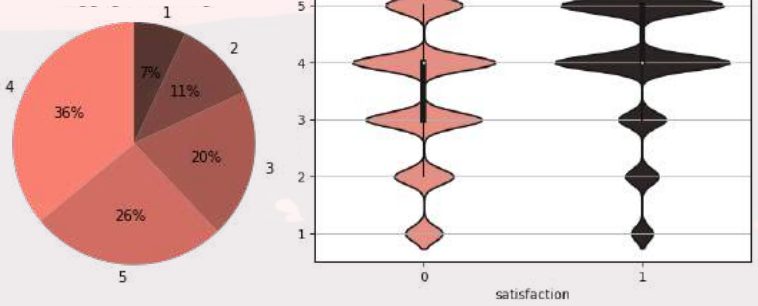
Departure/Arrival time convenient



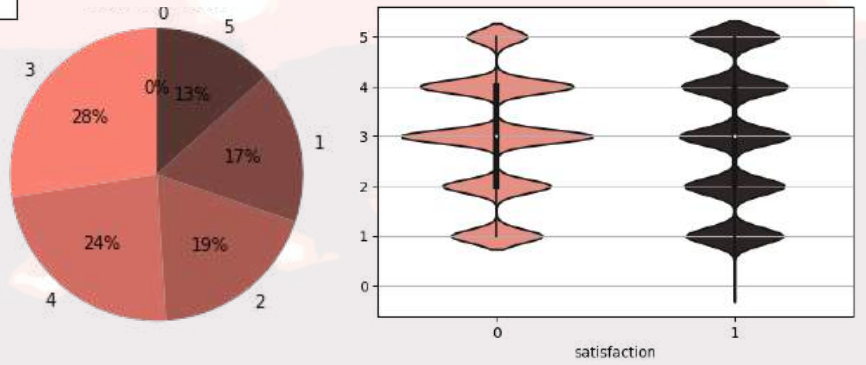
Online boarding



Baggage handling



Gate location

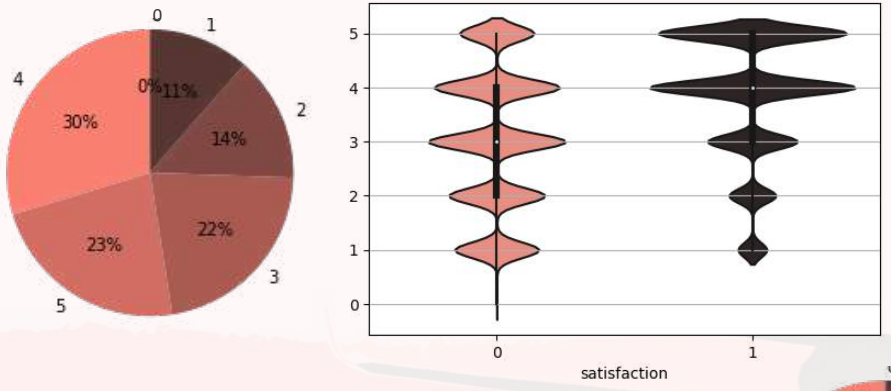


Exploratory data analysis

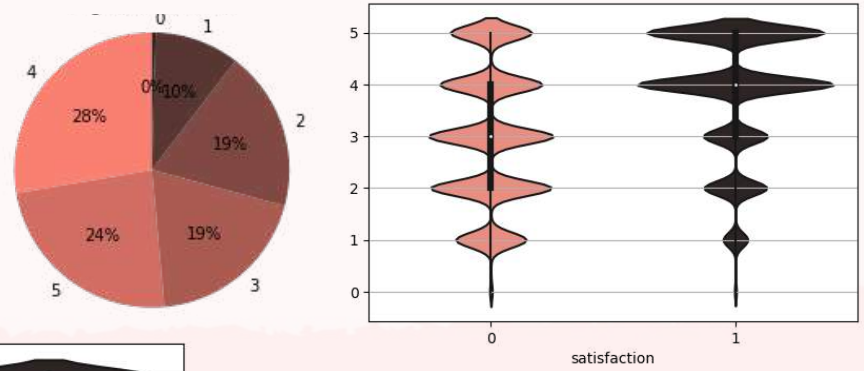
Hierarchical categorical features: services

Findings: Satisfaction comes more with better services.

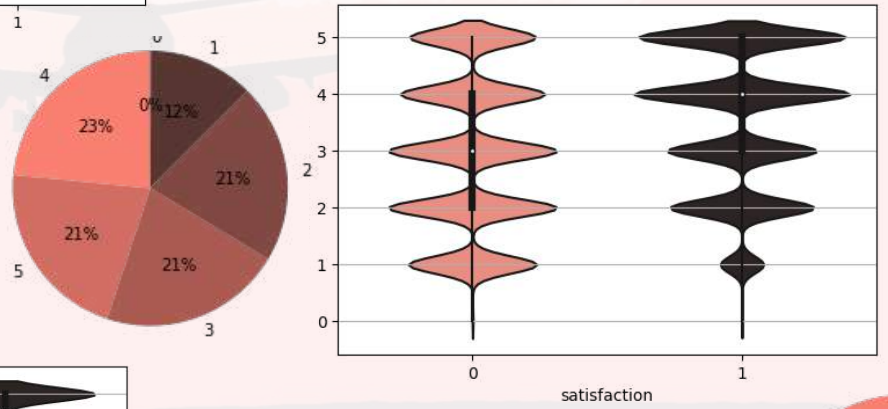
On-board service



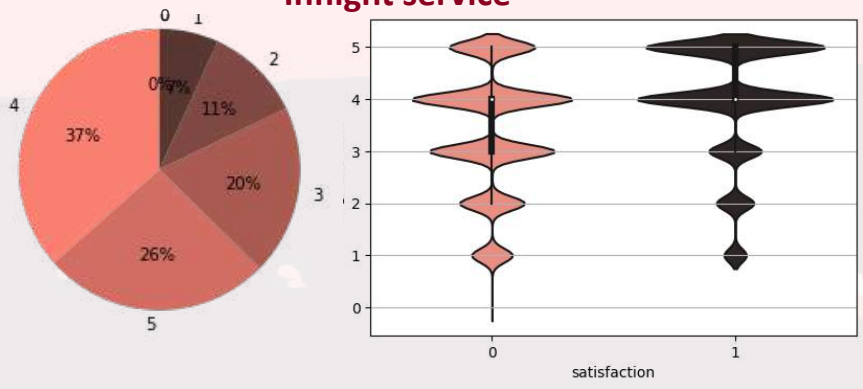
Leg room service



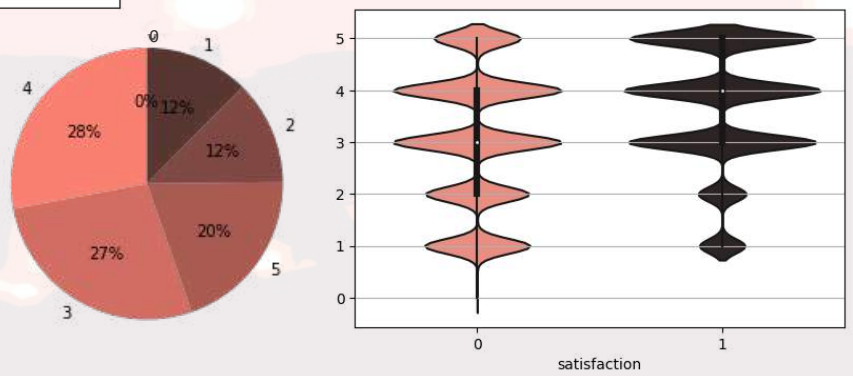
Food and drink



Inflight service



Checkin service



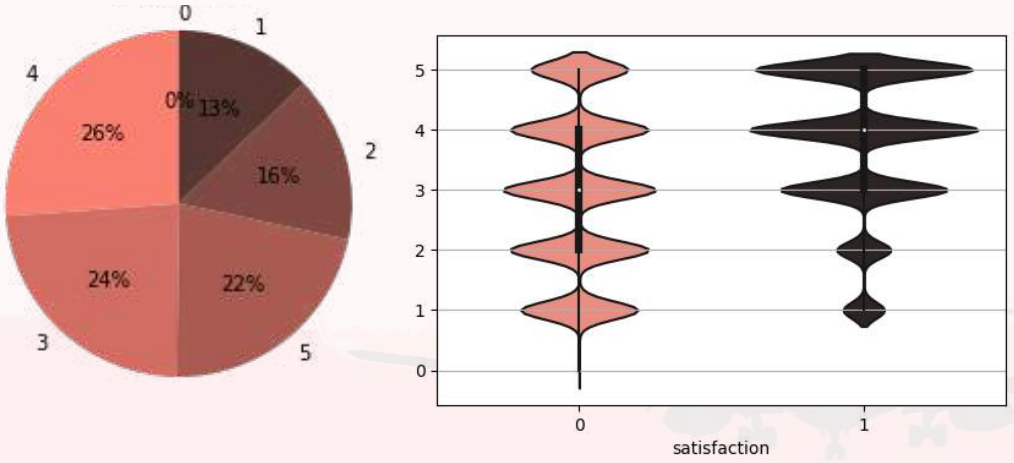
Exploratory data analysis

Hierarchical categorical features: inflight facilities

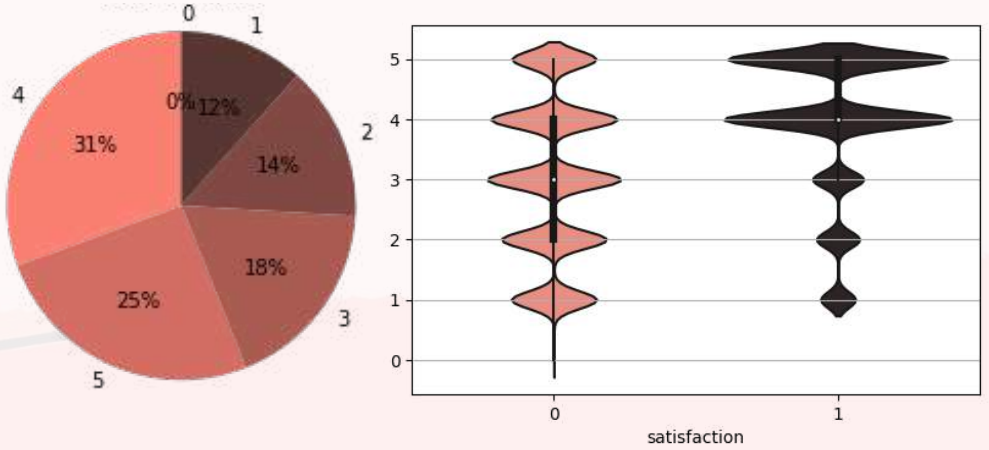
THE UNIVERSITY OF CHICAGO

Findings: satisfaction also comes more with better inflight facilities.

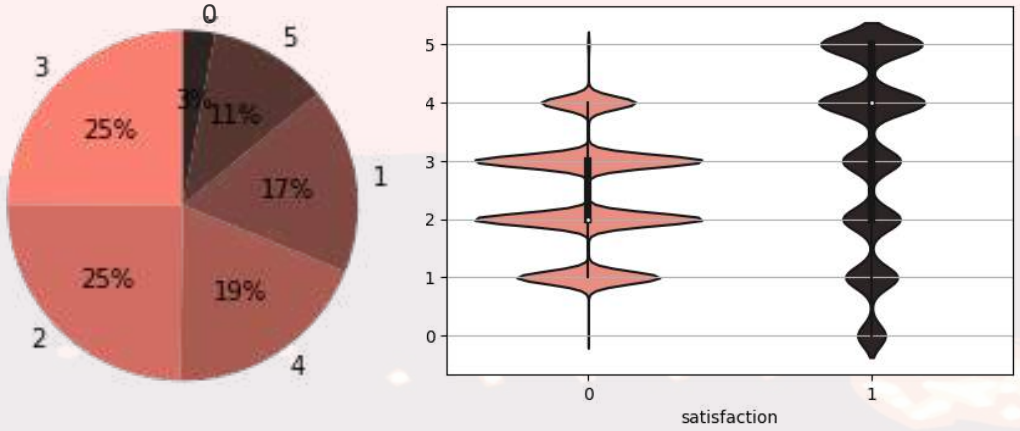
Cleanliness



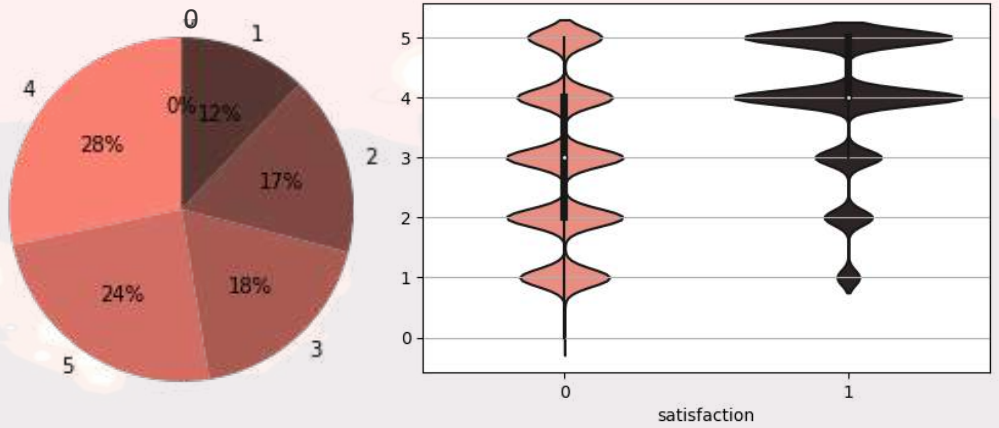
Seat comfort



Inflight wifi Service

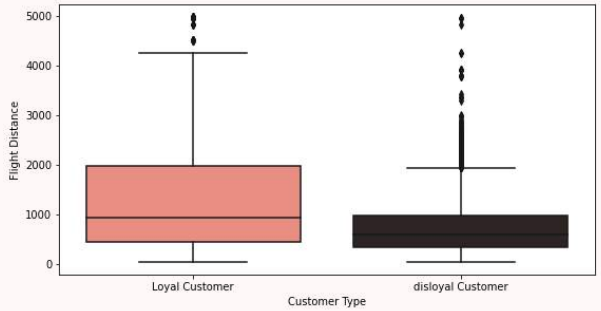


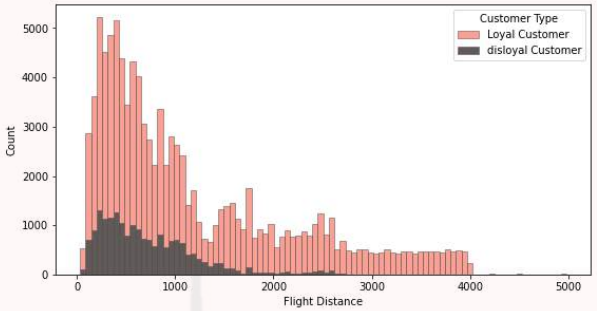
Inflight entertainment

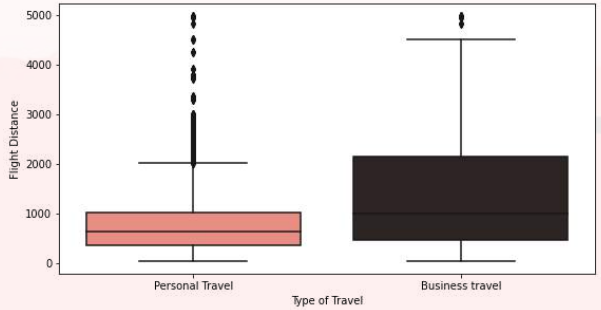


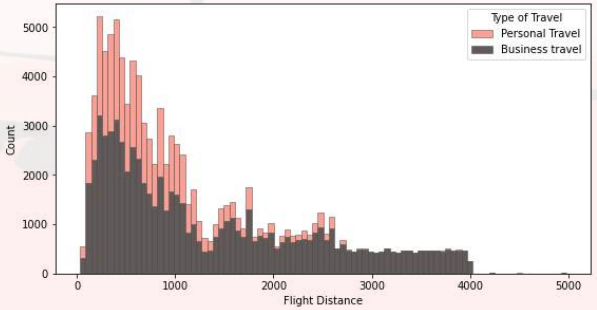
Exploratory data analysis

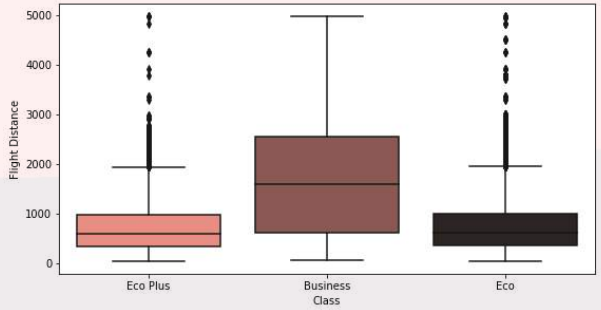
How flight distance correlate w/ selected features

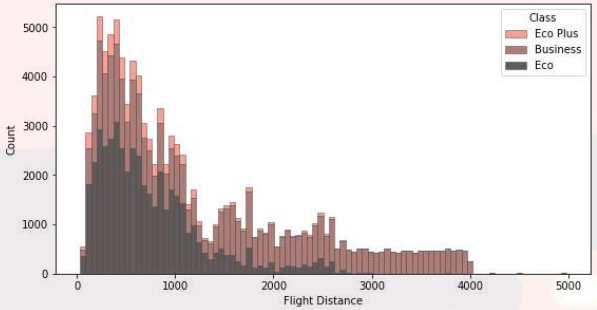










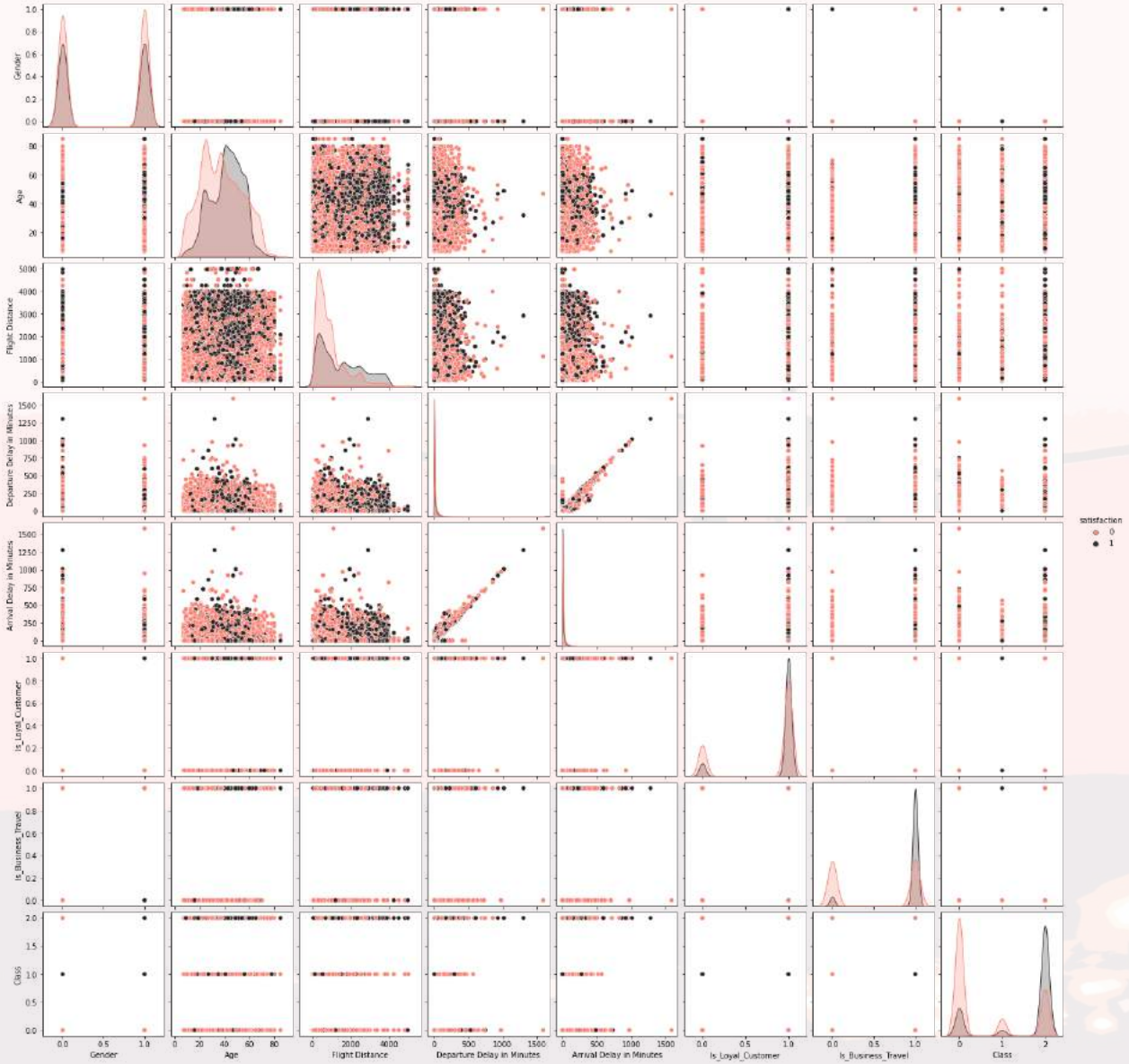


Findings

- Long distance flight passengers are prone to be loyal customers
- Long distance flight passengers generally travel for business purposes
- Long distance flight passengers tend to buy business class

Exploratory data analysis

Clustering – Try Out



Why: Generate different passenger groups to allows us make customized recommendations

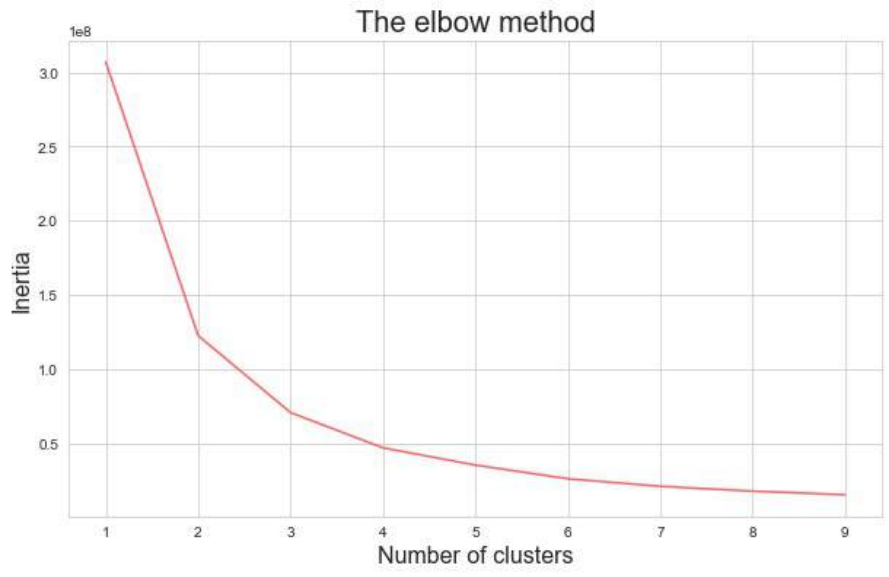
What: Clustering is an approach to generate labels and provide insight for feature engineering steps

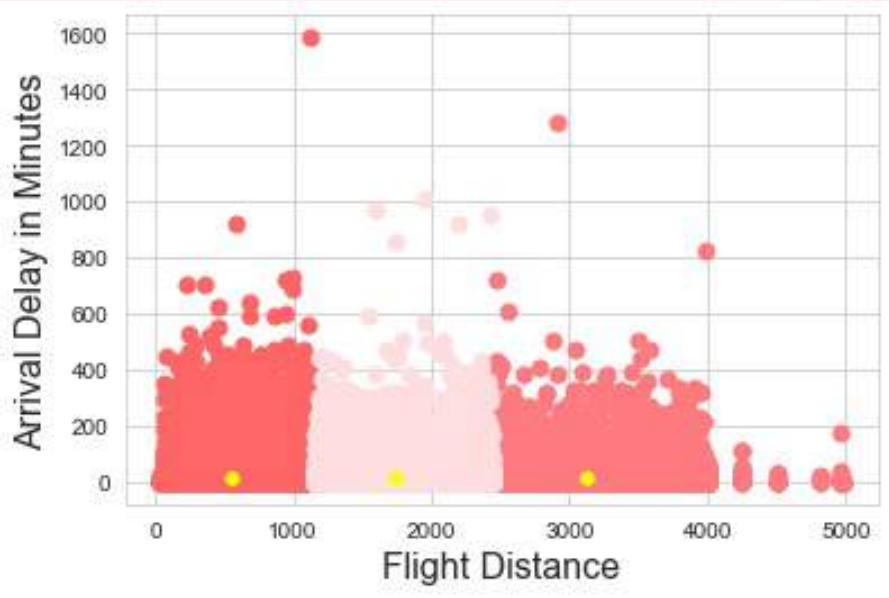
How: lessons learned in the process

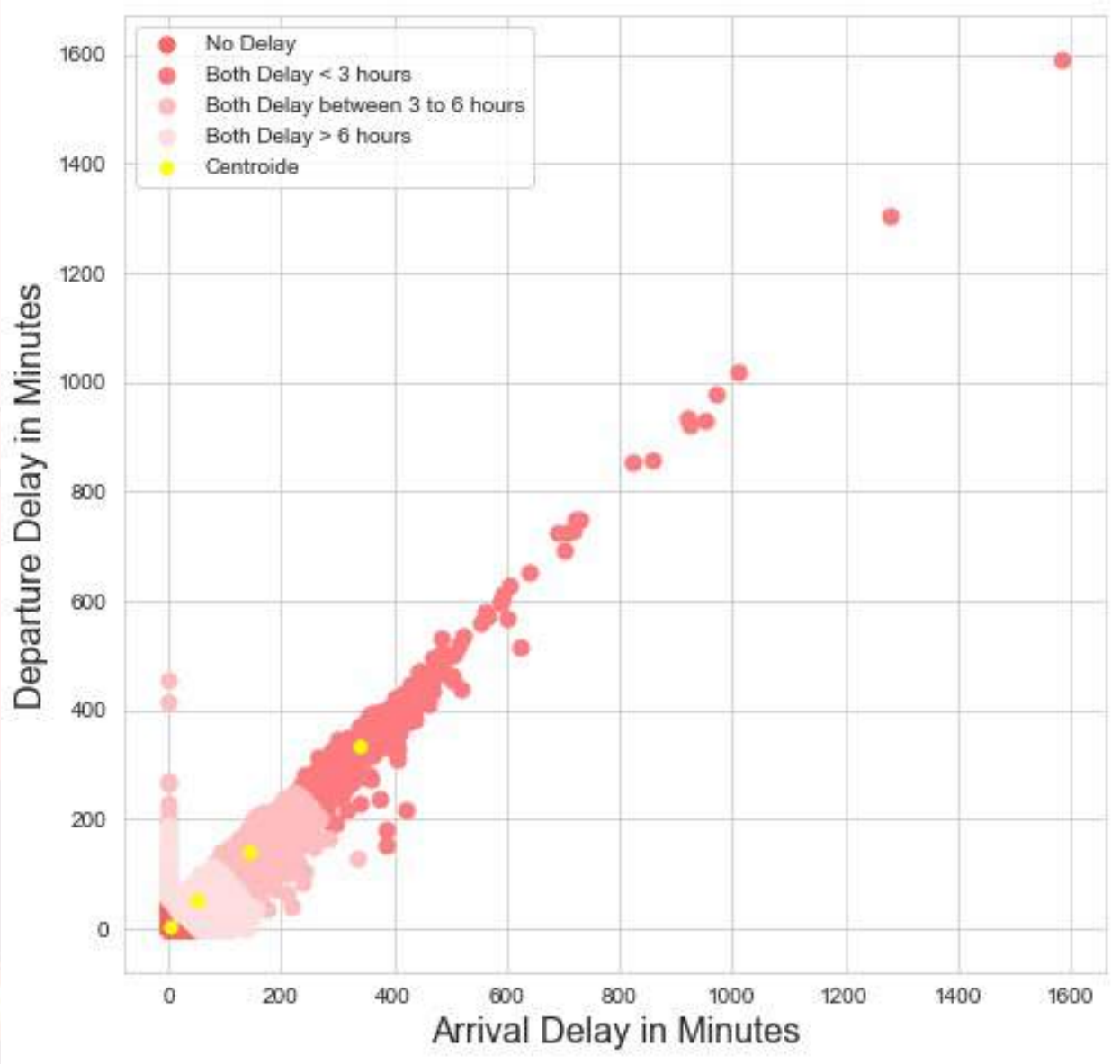
- The involvement of Binary or Categorical predictors would make clustering useless
- Using the combination of numerical predictors is the prerequisite
- It's necessary to check scatterplots for understanding the data before applyting clustering

Exploratory data analysis

Clustering – Result

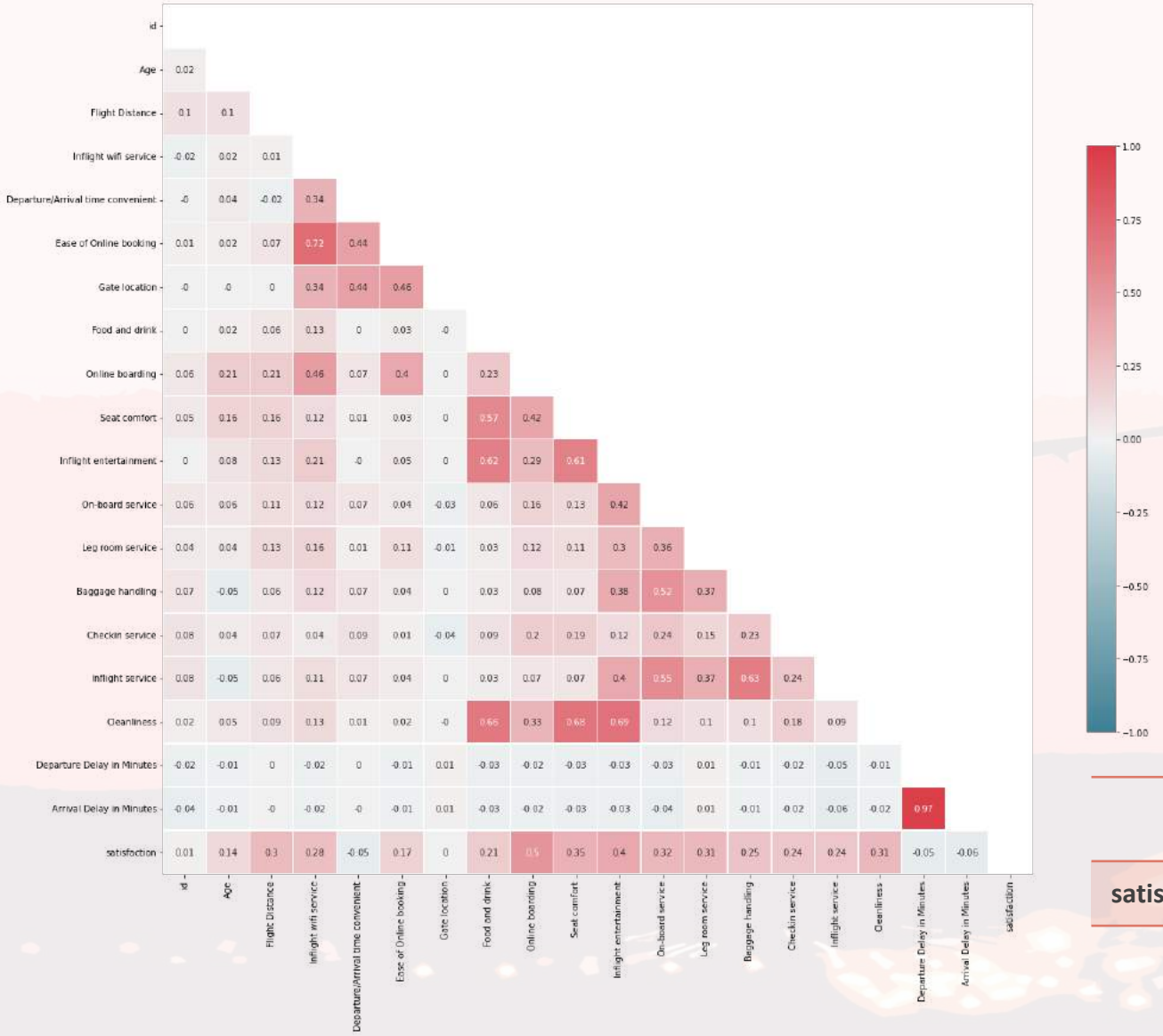






Feature Mining

Data cleaning & engineering



Cleaning

- Drop id, Gate location, Departure/Arrival time convenient due to low correlation
- Drop Arrival Delay in Minutes due to its high correlation with Departure Delay in Minutes and missing values

Engineering

Creating categorical features using clustering info & examining their correlation with satisfaction

	is_dep_delay	is_arr_delay	is_delay	delay_clusters	is_longdist	Flight Distance
satisfaction	-0.07	-0.09	-0.07	-0.07	0.28	0.3

Model Selection

Tree based & other models

THE UNIVERSITY OF CHICAGO

Tree Based Models: Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Light GBM, XGBoost

Other Models: Gaussian Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Stochastic Gradient Descent

Rank	Model	Accuracy	F1 Score
	**Light GBM**	**96.28**	**95.63**
2	Random Forest	96.24	95.6
3	XGBoost	96.16	95.5
4	Decision Tree	94.53	93.72
5	Gradient Boosting	94.01	93.01
6	KNN	93.6	92.44
7	AdaBoost	92.41	91.21
8	Logistic Regression	87.57	85.45
9	SGD	87.35	84.19
10	Linear SVM	87.3	85.21
11	Naive Bayes	85.29	82.97

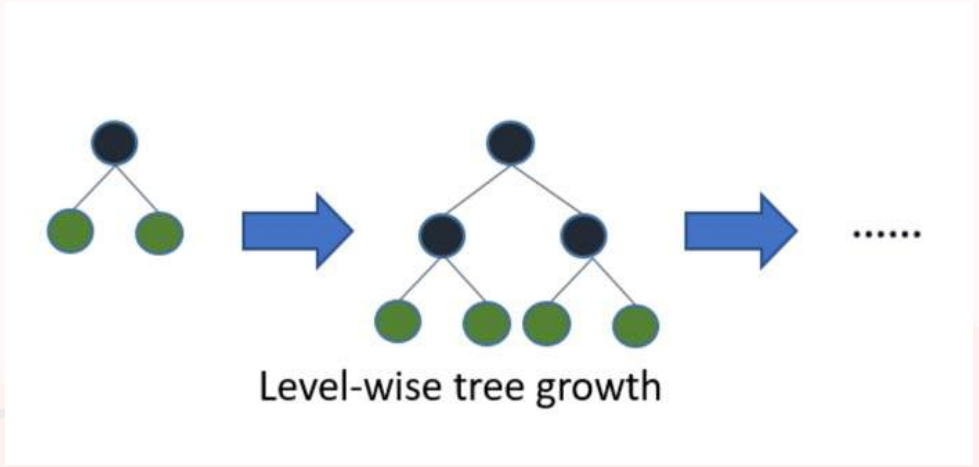
Light Gradient Boosting Machine (LightGBM)

Our final model

THE UNIVERSITY OF CHICAGO

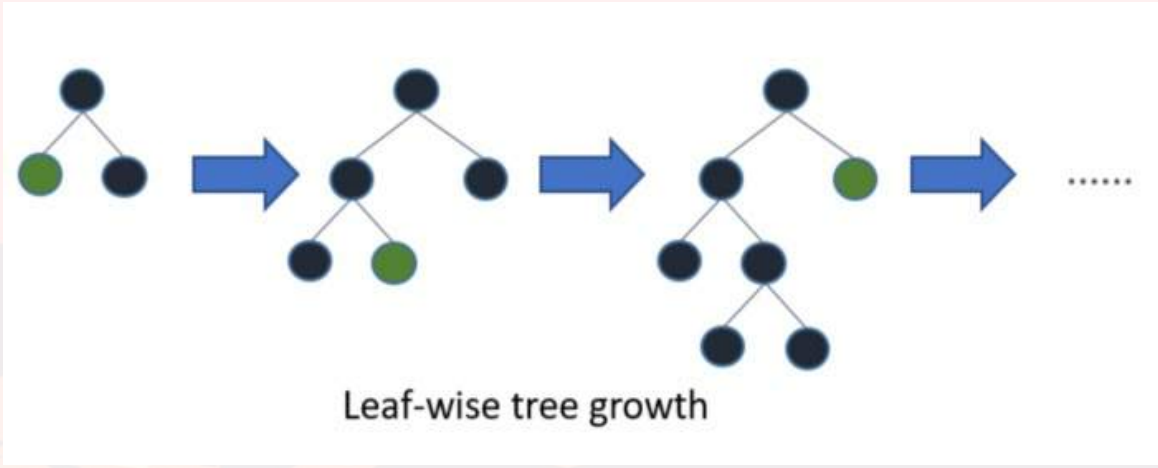
Supervised Machine Learning

- A gradient boosting framework that uses a tree-based algorithm
- Reduce complexity
  - Gradient based one side sampling
  - Exclusive feature bundling
  - Histogram based splits



Pros

- Faster training speed and higher efficiency
- Better accuracy
- Lower memory usage
- Support of parallel, distributed, and GPU learning
- Capable of handling large-scale data



Cons

- Overfitting
- Compatibility with Datasets

Success Metrics:

- Confusion Matrix
- Classification Report with Precision, Recall, F1 score

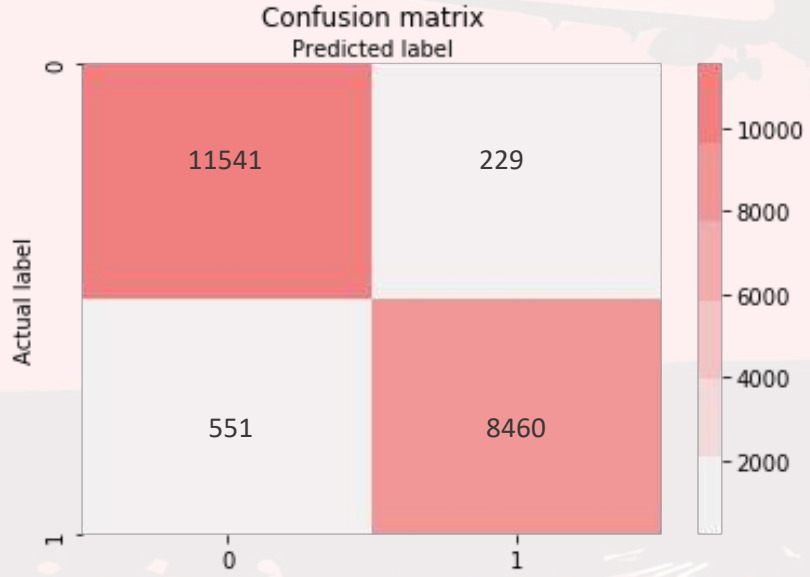
Model performance

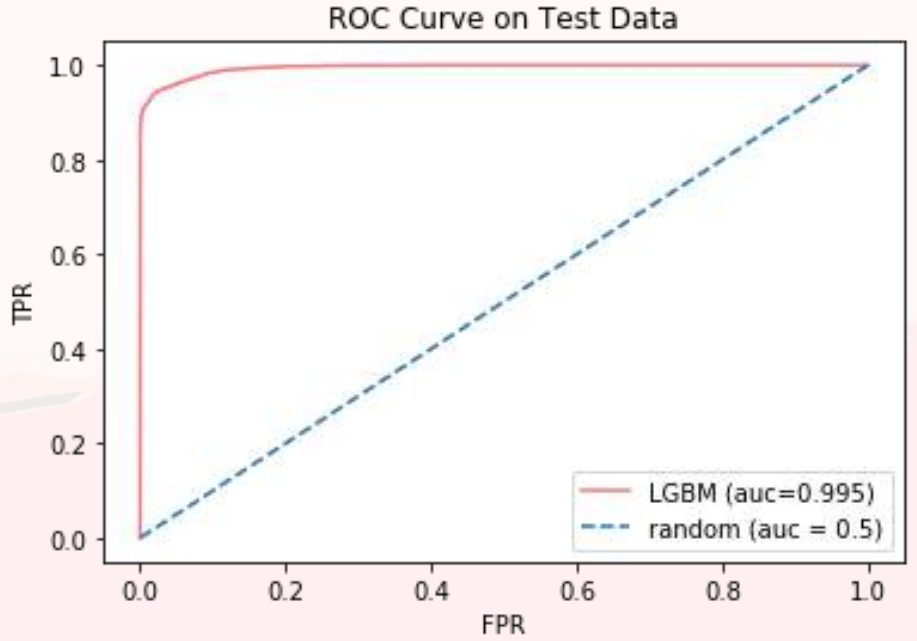
Success metrics w/ our final model: Light GBM

THE UNIVERSITY OF CHICAGO

Hyper-parameters tuning using cross-validation

- Maximum tree depth: 8
- Minimum children samples: 18
- Minimum children weight: 0.001
- Number of leaves: 40



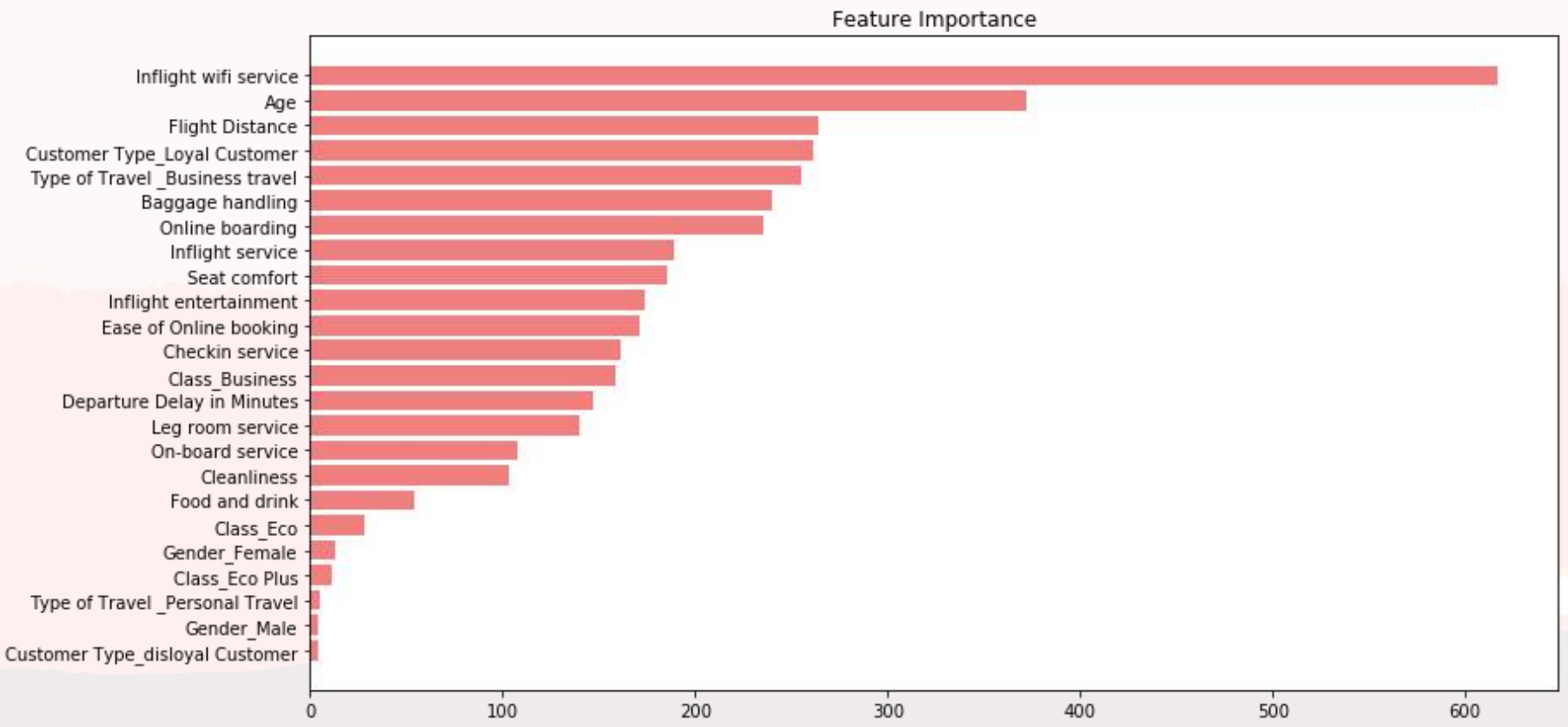


Classification Report

	Precision	Recall	F1 Score	Accuracy
0	0.95	0.98	0.97	
1	0.97	0.94	0.96	
overall				0.96

Model performance

Feature importance w/ Light GBM

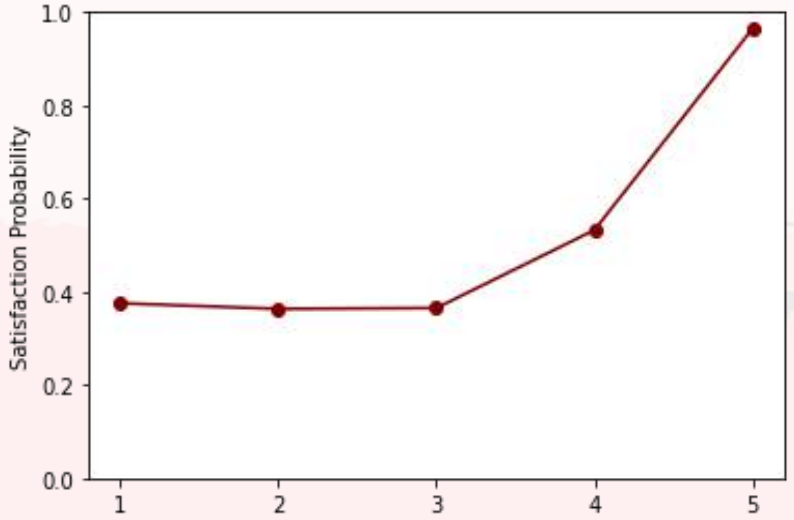


Top 5 Features	
1	Inflight wifi service
2	Age
3	Flight Distance
4	Customer Type
5	Type of Travel

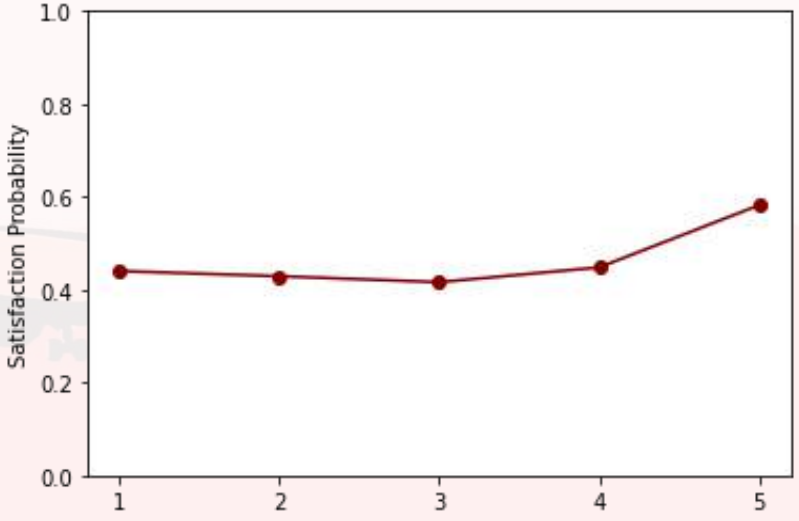
Top 5 upgradeable features	
1	Inflight wifi service
2	Baggage handling
3	Online boarding
4	Inflight service
5	Seat comfort

Feature predictions (Partial Dependence Plot)

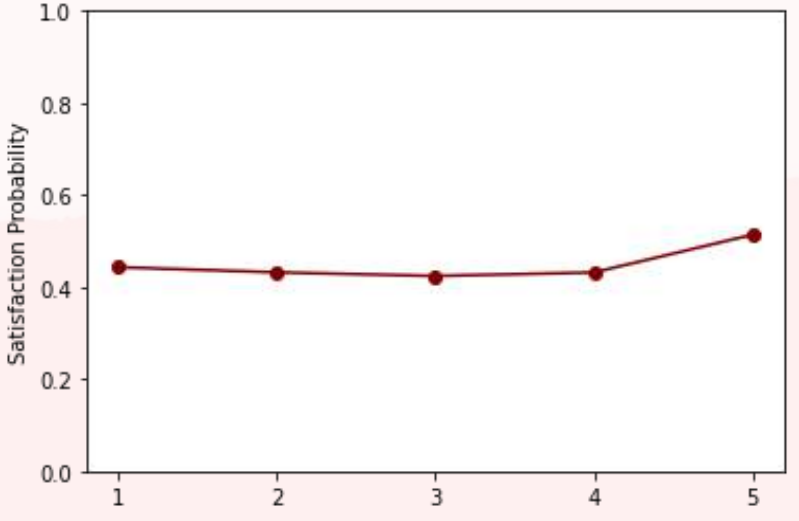
How much can changing feature values improve satisfaction in probability?



Inflight wifi Service



Online boarding



Seat comfort

Improve by 0.6

Improve by 0.14

Improve by 0.07

Interpretation

Differentiating customers

We intend to generate labels for passengers based on objective features

Differentiation allows us to make customized recommendations:

- **Pools**
  - Gender, Age, Flight Distance, Loyalty, Type of Travel, Class
- **Selection**
  - Loyalty
  - Type of Travel
  - Class

- Example:

Loyalty \ Type of Travel	Business	Non-Business
True	**Frequent Passenger** - High value customer, given we assume business travels and loyalty passengers would use our service most frequently	**Loyal Traveler** - Our main supporters, should constantly observe their feedback to make sure no churn
False	**Frequent Newcomer** - Potential high value customers, should attract them	**Mobile Traveler** - Travelers don't care about which airline, no need further operation

Interpretation

Loyalty/Class

THE UNIVERSITY OF CHICAGO

Customer Type		Inflight wifi service	Baggage handling	Online boarding	Inflight service	Seat comfort	Satisfaction
Economy	non-Loyal	2.58	3.37	2.59	3.38	3	0.14
	Loyal	2.7	3.47	2.88	3.49	3.18	0.2
Economy Plus	non-Loyal	2.46	3.14	2.47	3.15	3.02	0.08
	Loyal	2.8	3.39	2.93	3.41	3.2	0.26
Business	non-loyal	2.93	4.22	2.91	4.22	**2.99**	0.4
	Loyal	**2.75**	3.78	3.86	3.78	3.9	0.75

Findings

- Loyal Business Class passengers aren't satisfied with **inflight wifi service**
- Non-Loyal Business Class passengers aren't satisfied with **seat comfort**
- Nearly no group is satisfied with the existing **inflight wifi service**

Interpretation

Type of Travel/Class

THE UNIVERSITY OF CHICAGO

Customer Type		Inflight wifi service	Baggage handling	Online boarding	Inflight service	Seat comfort	Satisfaction
non-Business Travel	Economy	2.52	3.6	2.77	3.62	3.17	0.1
	Economy Plus	2.5	3.57	2.76	3.62	3.21	0.09
	Business Class	2.55	3.34	3.24	3.33	3.46	0.12
Business Travel	Economy	2.88	3.26	2.87	3.26	3.1	0.3
	Economy Plus	3.01	3.17	3.01	3.17	3.16	0.39
	Business Class	**2.79**	3.87	3.74	3.87	3.77	0.72

Findings

- Business Travel who also booked business class might potentially care about working in flights, poor **inflight wifi service** would be a huge CON for their overall experience

Interpretation

Type of Travel/Loyalty

THE UNIVERSITY OF CHICAGO

Customer Type		Inflight wifi service	Baggage handling	Online boarding	Inflight service	Seat comfort	Satisfaction
non-Business Travel	non-Loyal Customer	2.63	3.73	**2.48**	3.78	2.98	0.16
	Loyal Customer	2.52	3.58	**2.8**	3.6	3.2	0.1
Business Travel	non-Loyal Customer	2.71	3.69	**2.71**	3.7	2.99	0.24
	Loyal Customer	2.86	3.64	3.72	3.64	3.75	0.71

Findings

- Baggage Handling and Inflight service are exceptional
- **Online Boarding** is not satisfying the majority groups

Recommendation

Potential area of improvement

THE UNIVERSITY OF CHICAGO

Inflight Wifi service

Seat Comfort

Online Boarding

Recommendation

Analysis on Inflight Wifi Service

Revenue

- Approx. 5% of passengers would use
- Normal rates: hour-pass $10, full-use $20
- Estimate revenue per flight: $10 * 200 * 5% ≈ $100

Cost

- Installation: $200,000 ~ $300,000 per plane
- Recurring Service Cost: $4345 ~ $9995 /month
- Labor cost: 4-6 mechanics * 72 hours * $30/hour (for 737/A320)

Payback Period: Approx. **41 months**

Recommendation

Analysis on Seat Comfort

Production Cost

- **$3000 ~ $5000** per economy class seat; **$10000** for long flights
- **$30000 ~ $80000** per business class seat

* World most popular plane (Boeing 737-800: 189 economy seats occupancy)

Estimated Production Cost: **$1,890,000**

Design and Installation

- New seat designment is the stage that takes most of the time
- Hard for groundbreaking improvement

Recommendation

Analysis on Online Boarding

Software Development is the key

Time:

- 30 days for building up an application from scratch
- 2 to 3 weeks at most for version update

Labor:

- A team of 5 including project manager and CS engineers

Cost:

- Software Development:

  5 people * 30 days * 8 hours/day * $50/hour = $60,000

Total Cost: **$100,000** (development + maintenance)

Recommendation

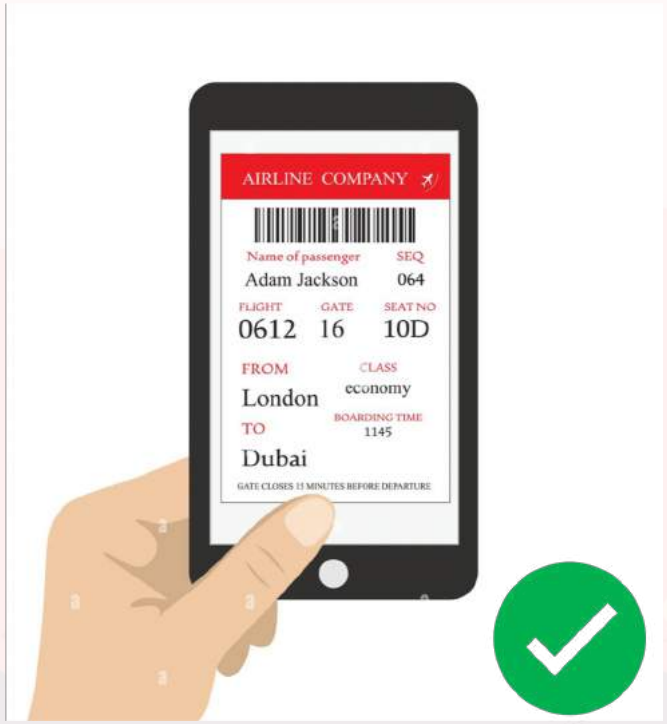
Summary

THE UNIVERSITY OF CHICAGO

Inflight Wifi Service

Seat Comfort

Online Boarding




Q&A