

Happier Passengers

An Analysis of an Airline Passenger Survey

Problem Statement and Goal

Problem Statement:

- How to increase average passenger satisfaction given finite resources?

An Approach:

- a) Quantify and rank the impact of various factors of flight quality
- b) Based on the rankings and findings, tailor the flight process

Goal and Scope

- Implement a)

Dataset Description

Survey of US airline passengers taken from Kaggle ([link](#))

- **Sample size of ~130K**
- **Variables:**
 - **Continuous Variables:** Age, Flight Distance, Delays
 - **Categorical Variables (String):** Gender, Customer Type, Type of Travel, Class, satisfaction [target]
 - **Categorical Variables (1:Worst..5:Best; 0:N/A)**

Inflight wifi service | Departure/Arrival Time Convenient |

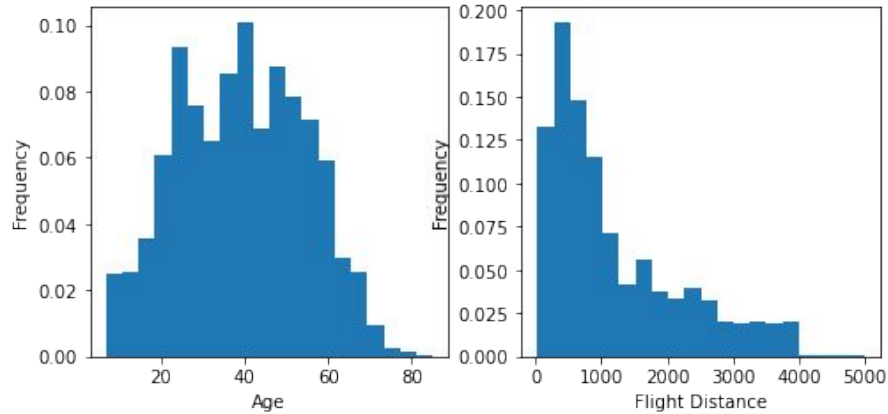
Ease of Online Booking | Gate location | Food and drink |

Seat Comfort | Online boarding | Inflight entertainment |

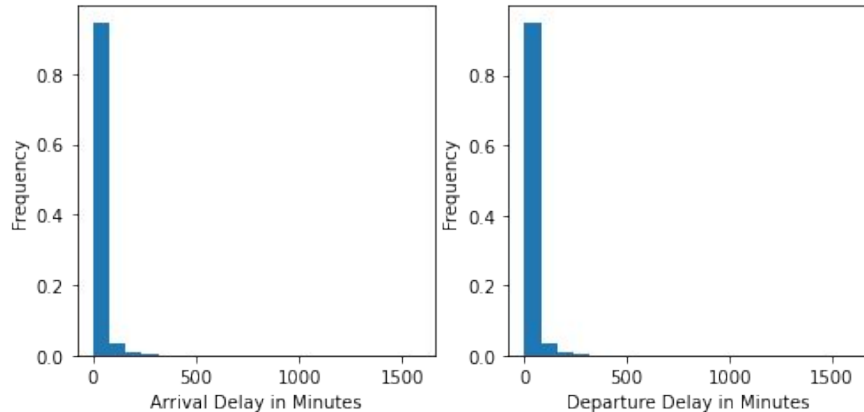
On-board service | Leg room service | Baggage handling |

Checkin service | Inflight service | Cleanliness |

EDA - Continuous Variables

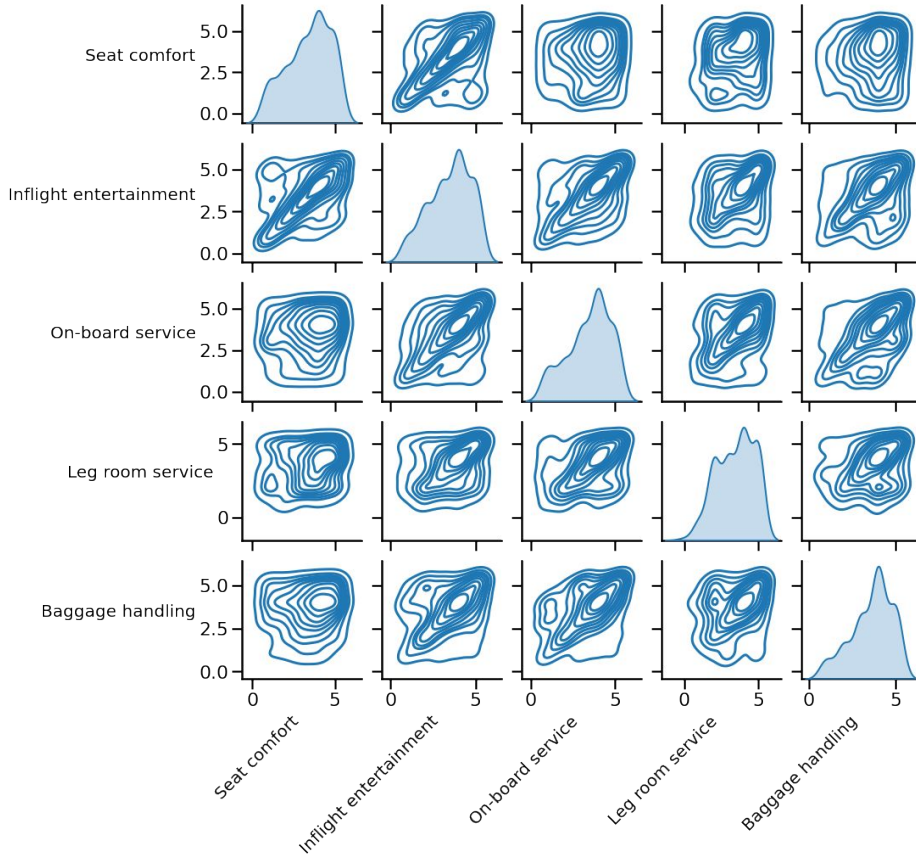


>> Flight Distance can be split: [<900 , ≥ 900]?



>> Mass around 0 with large outliers

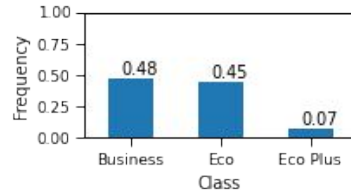
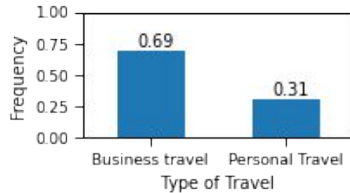
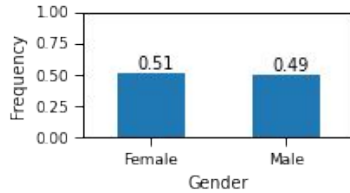
EDA - Categorical (1..5)



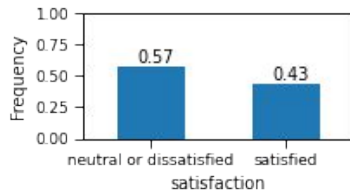
-Most distributions unimodal

>> Bimodal may point to information to be mined

EDA - Categorical String Variables



>> Business sample ~ Economy sample. Why?



>> 'satisfaction' [target variable] looks balanced

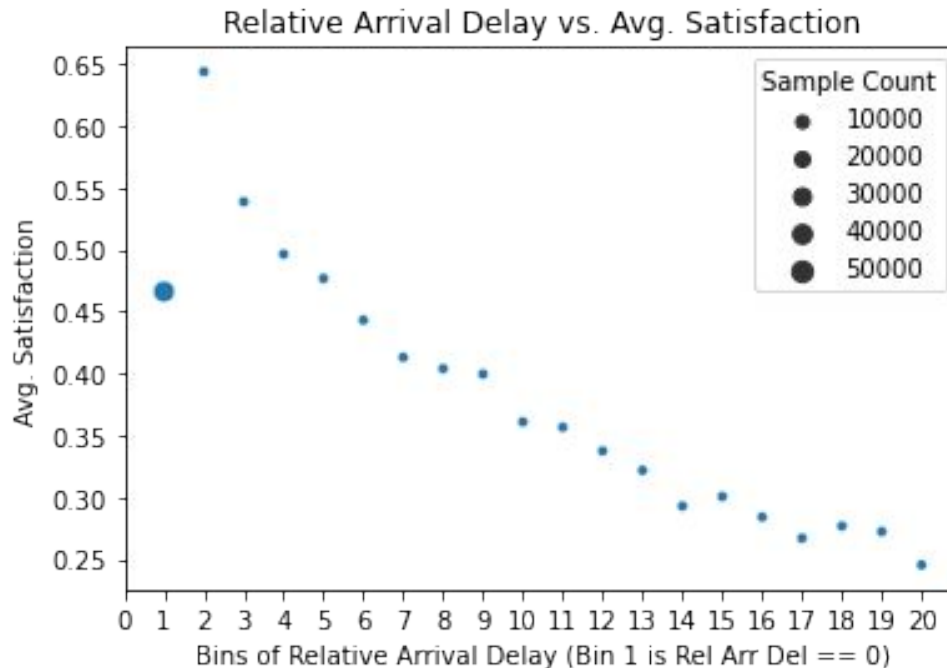
Data Wrangling

- **Encode categorical string variables**
- **Drop NA's from dataset**
- **Drop 0's from survey questions**

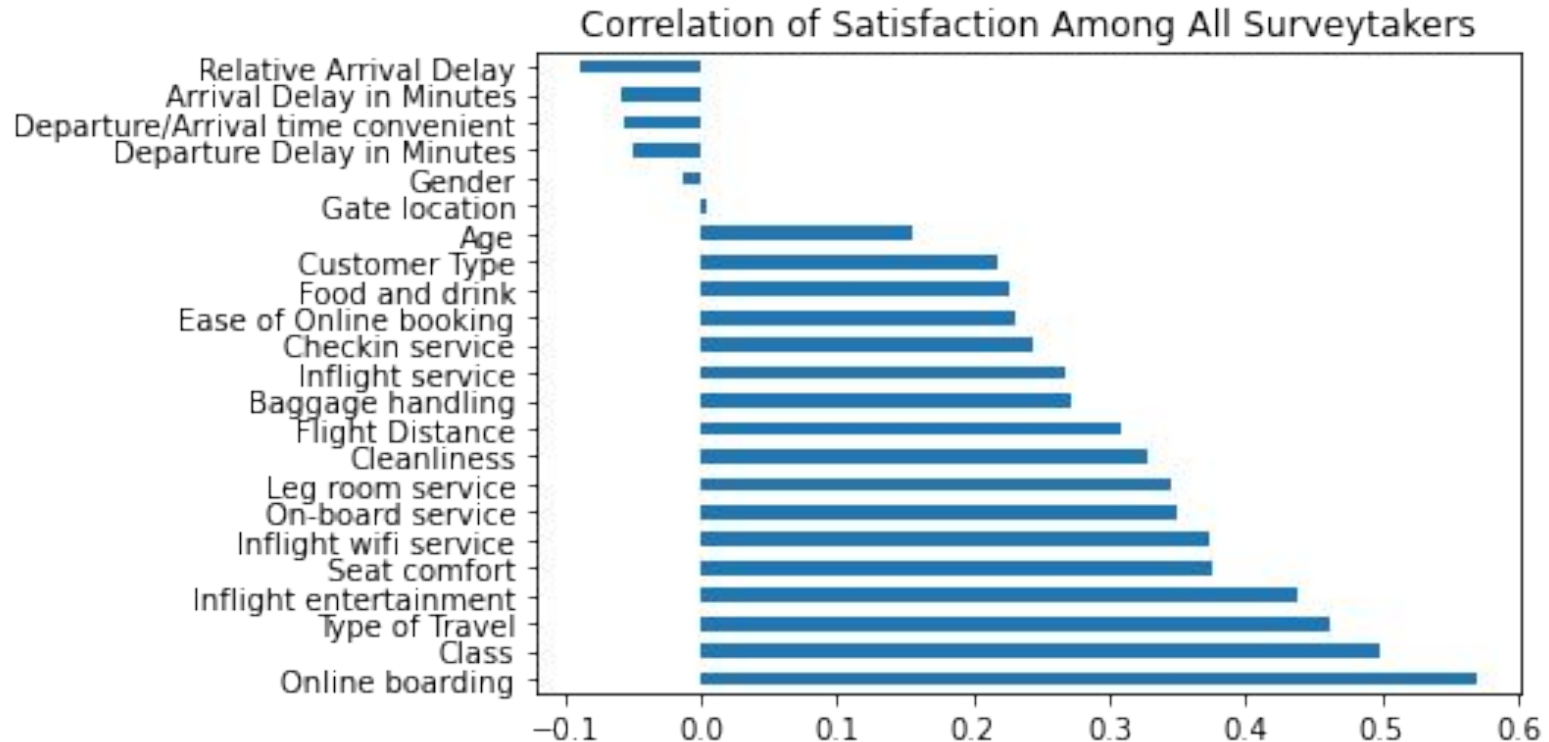
	Training Size	Test Size
Before Drop	103,904	25,976
After NA Drop	103,594	25,893
After 0 Drop	95,415	23,789

Feature Engineering

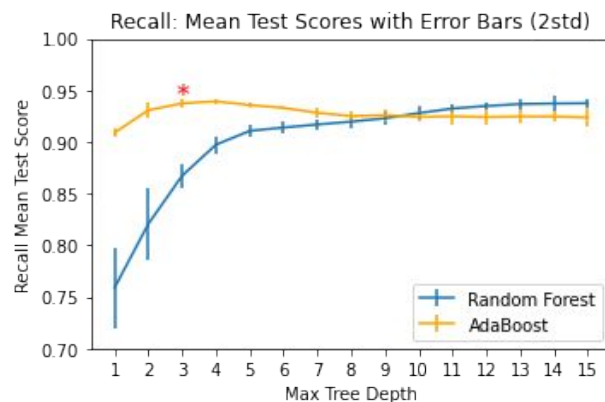
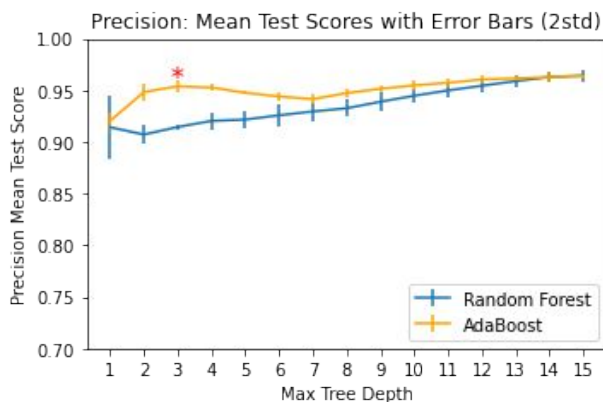
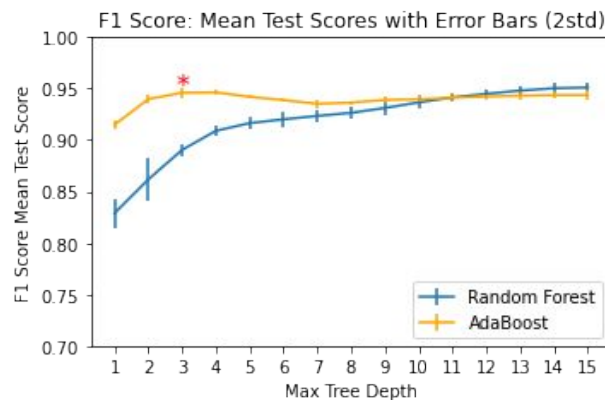
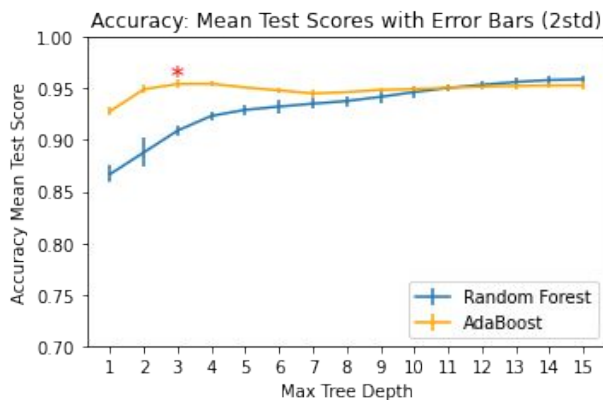
$$\begin{array}{lcl} \textit{Relative Arrival Delay} & = & \textit{Arrival Delay in Minutes} / \textit{Flight Distance} \\ (\text{corr w/ target} = -0.088) & & (\text{corr w/ target} = -0.058) \end{array}$$



Modeling - Input Correlations with 'satisfaction'



Modeling - Random Forest vs AdaBoost

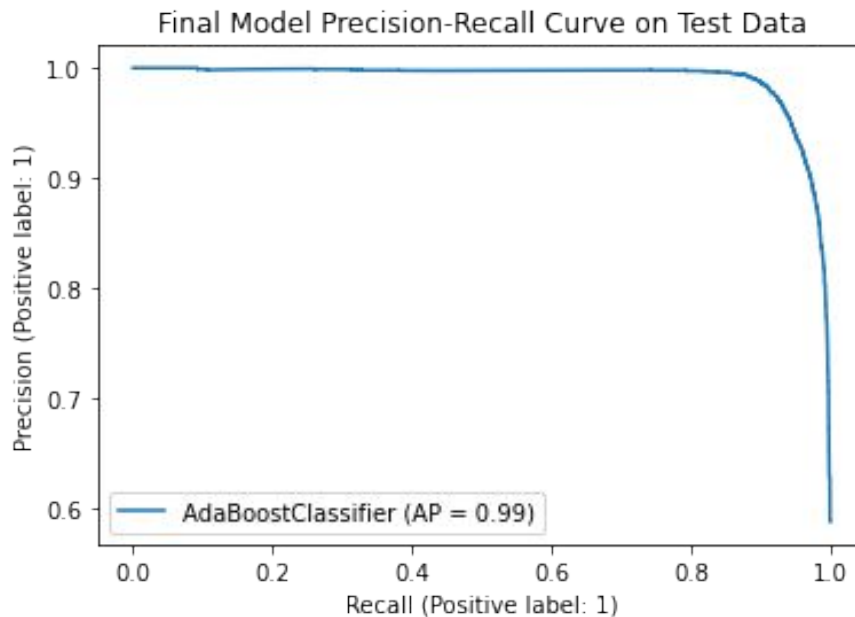


Modeling - Final Model (Scores)

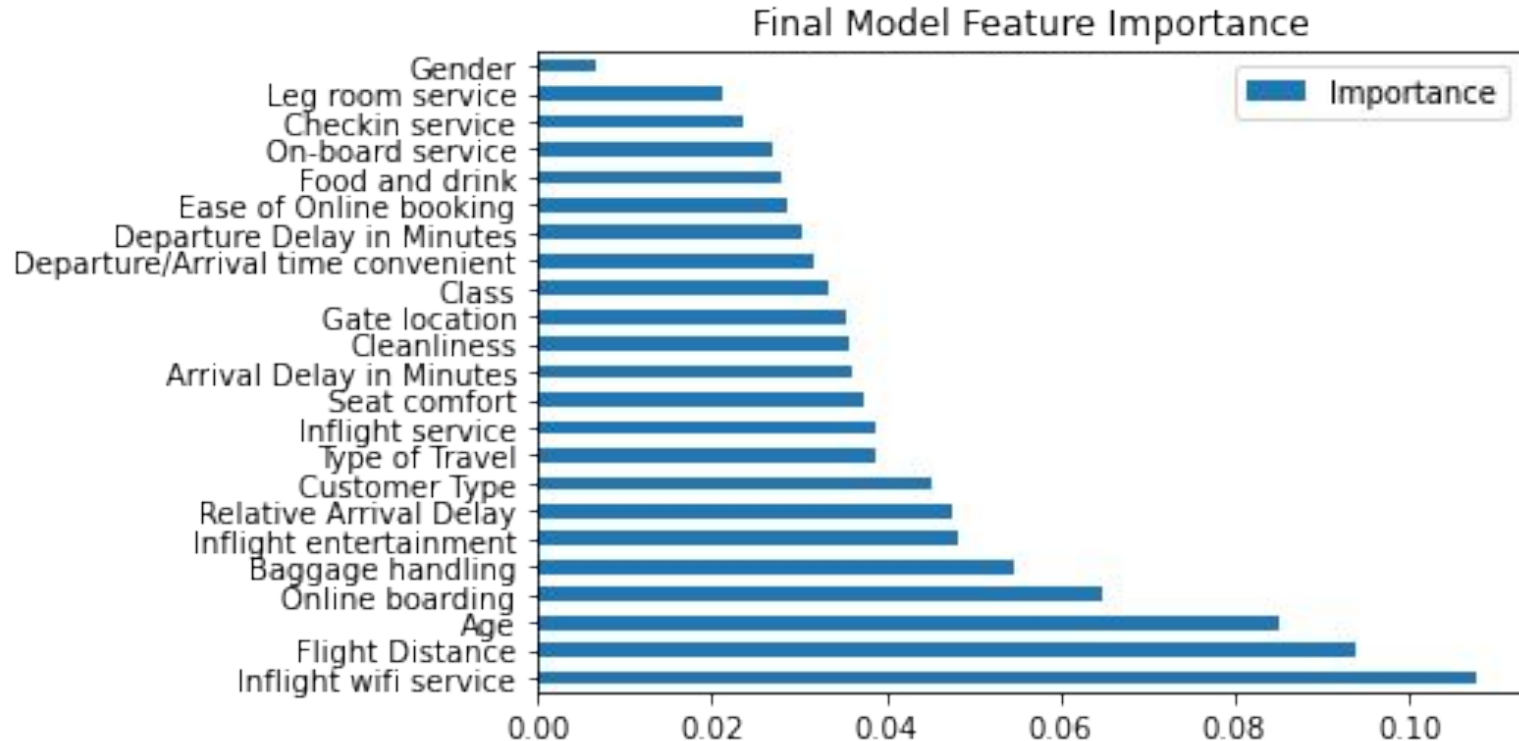
Final Model: AdaBoost with Max Tree Depth of 3 and 100 estimators

-Trained on all training data

Final Model (AdaBoost) Scores on Test Data	
Accuracy	0.956 <input type="button" value="▼"/>
F1 Score	0.949
Precision	0.956
Recall	0.941



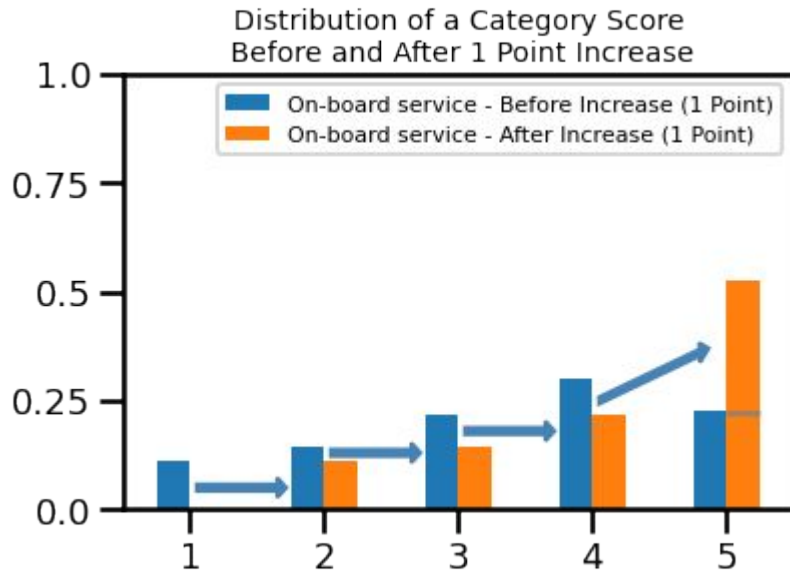
Modeling - Final Model (Feature Importance)



Sensitivity Analysis - Distribution Shift

Key Question:

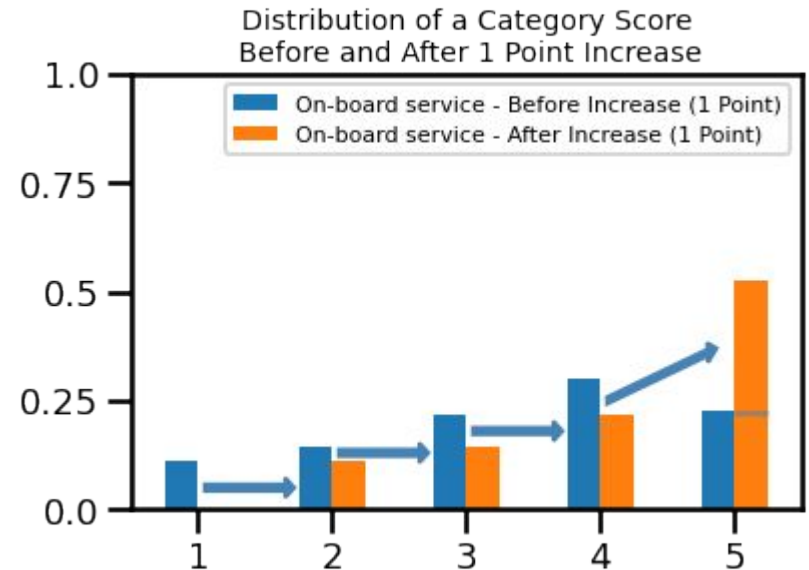
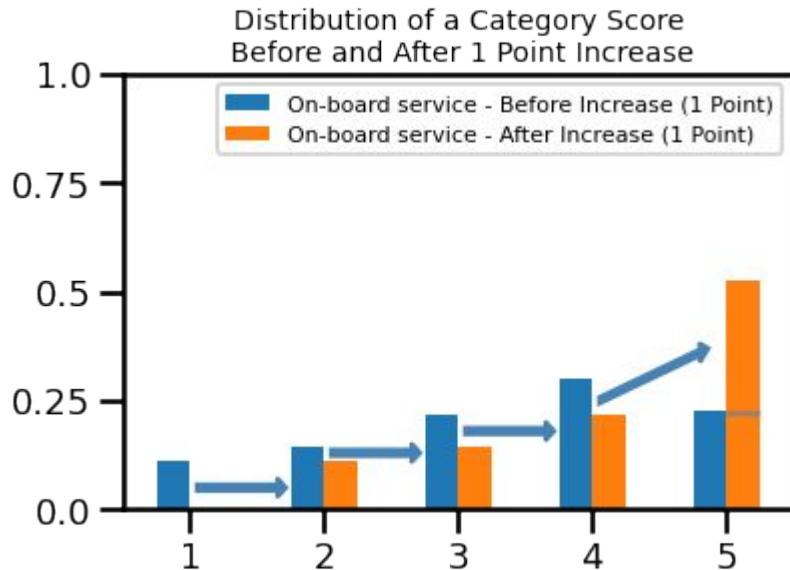
If all the passengers who responded to, say, 'On-board service' with a 3 (out of 5) were to increase their scores to 4, how would this group's average satisfaction change?



Sensitivity Analysis - Distribution Shift

Key Question:

If all the passengers who responded to, say, 'On-board service' with a 3 (out of 5) were to increase their scores to 4, how would this group's average satisfaction change?

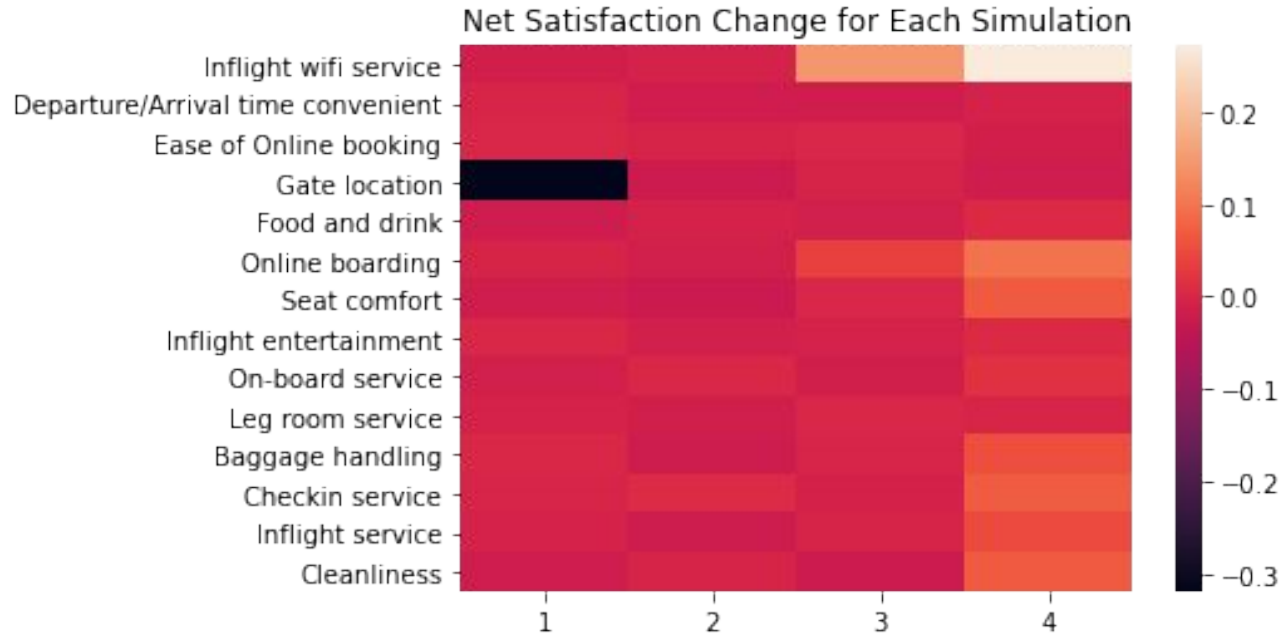


Sensitivity Analysis - Avg. 'satisfaction': Before and After

Average Satisfaction per Category and Score					
	1	2	3	4	5
Inflight wifi service	36.4%	25.3%	26.8%	61.1%	98.7%
Departure/Arrival time convenient	50.2%	44.2%	44.2%	38.3%	41.3%
Ease of Online booking	39.5%	31.2%	32.6%	53.3%	71.0%
Gate location	50.6%	46.5%	33.4%	38.9%	55.1%

Average Satisfaction after Scores increase by 1					
	1->2	2->3	3->4	4->5	N/A
Inflight wifi service	35.0%	24.8%	41.5%	88.3%	NaN
Departure/Arrival time convenient	50.1%	42.4%	42.7%	37.7%	NaN
Ease of Online booking	39.5%	30.8%	32.6%	51.9%	NaN
Gate location	18.6%	44.2%	33.0%	37.0%	NaN

Sensitivity Analysis - Net Satisfaction Change



Sensitivity Analysis - Impact of Distribution Shift

Impact of 1 Point Increase on Average Satisfaction (Sorted)

	Impact of 1 Point Increase	Baseline Satisfaction	Satisfaction After 1 Point Increase
Inflight wifi service	8.8%	43.0%	51.8%
Online boarding	3.5%	43.0%	46.5%
Checkin service	1.8%	43.0%	44.8%
Baggage handling	1.8%	43.0%	44.8%
Seat comfort	1.6%	43.0%	44.6%
Inflight service	1.5%	43.0%	44.5%
Cleanliness	0.9%	43.0%	44.0%
On-board service	0.2%	43.0%	43.2%
Inflight entertainment	-0.1%	43.0%	42.9%
Leg room service	-0.3%	43.0%	42.7%
Ease of Online booking	-0.4%	43.0%	42.7%
Food and drink	-0.4%	43.0%	42.6%
Departure/Arrival time convenient	-0.8%	43.0%	42.3%
Gate location	-6.3%	43.0%	36.7%

Recommendations 1 (Based on Simulation Results)

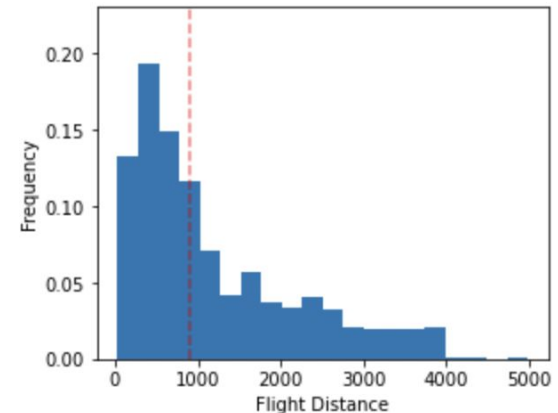
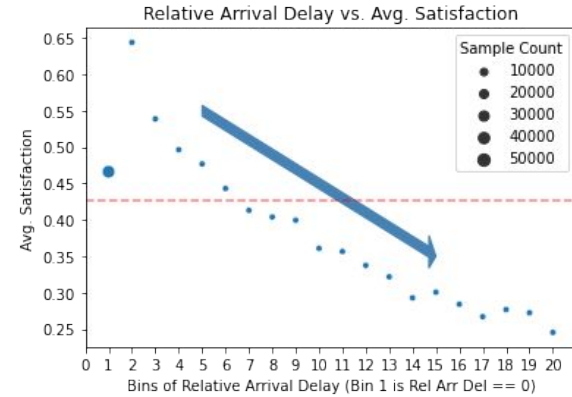
- **At first sight - Implement the following changes**
 - **Improve 'Inflight wifi service'**
 - **Improve 'online boarding'**
- **Before implementation however...**
 - **Evaluate cost per satisfaction change**
 - **Run small-scale experiments to confirm validity**

Impact of 1 Point Increase on Average Satisfaction (Sorted)

	Impact of 1 Point Increase	Baseline Satisfaction	Satisfaction After 1 Point Increase
Inflight wifi service	8.8%	43.0%	51.8%
Online boarding	3.5%	43.0%	46.5%
Checkin service	1.8%	43.0%	44.8%
Baggage handling	1.8%	43.0%	44.8%
Seat comfort	1.6%	43.0%	44.6%
Inflight service	1.5%	43.0%	44.5%
Cleanliness	0.9%	43.0%	44.0%
On-board service	0.2%	43.0%	43.2%
Inflight entertainment	-0.1%	43.0%	42.9%
Leg room service	-0.3%	43.0%	42.7%
Ease of Online booking	-0.4%	43.0%	42.7%
Food and drink	-0.4%	43.0%	42.6%
Departure/Arrival time convenient	-0.8%	43.0%	42.3%
Gate location	-6.3%	43.0%	36.7%

Recommendations 2 (Based on EDA)

- Investigate 'Relative Arrival Delay'
 - Monotonic decrease over late bins
 - Bins above and below baseline
(how/why do they differ?)
- Create sub-models by 'Flight Distance'
 - Long Distance vs Short Distance (red line=900)
 - Note the its high ranking in feature importance list



Some Future Research Directions

- **Design field experiments to check analysis validity**
- **Investigate causes of 'Relative Arrival Delay'**
- **Analyze short- and long-distance flights separately**
- **Code the infrastructure for analysts to enable more granular simulations**

Thank you & Open Floor for Discussion

Problem Statement:

- How to increase average passenger satisfaction given finite resources?

An Approach:

a) Quantify and rank the impact of various factors of flight quality

b) Based on the rankings and findings, tailor the flight process

Goal and Scope

- Implement a)

Impact of 1 Point Increase on Average Satisfaction (Sorted)

	Impact of 1 Point Increase	Baseline Satisfaction	Satisfaction After 1 Point Increase
Inflight wifi service	8.8%	43.0%	51.8%
Online boarding	3.5%	43.0%	46.5%
Checkin service	1.8%	43.0%	44.8%
Baggage handling	1.8%	43.0%	44.8%
Seat comfort	1.6%	43.0%	44.6%
Inflight service	1.5%	43.0%	44.5%
Cleanliness	0.9%	43.0%	44.0%
On-board service	0.2%	43.0%	43.2%
Inflight entertainment	-0.1%	43.0%	42.9%
Leg room service	-0.3%	43.0%	42.7%
Ease of Online booking	-0.4%	43.0%	42.7%
Food and drink	-0.4%	43.0%	42.6%
Departure/Arrival time convenient	-0.8%	43.0%	42.3%
Gate location	-6.3%	43.0%	36.7%