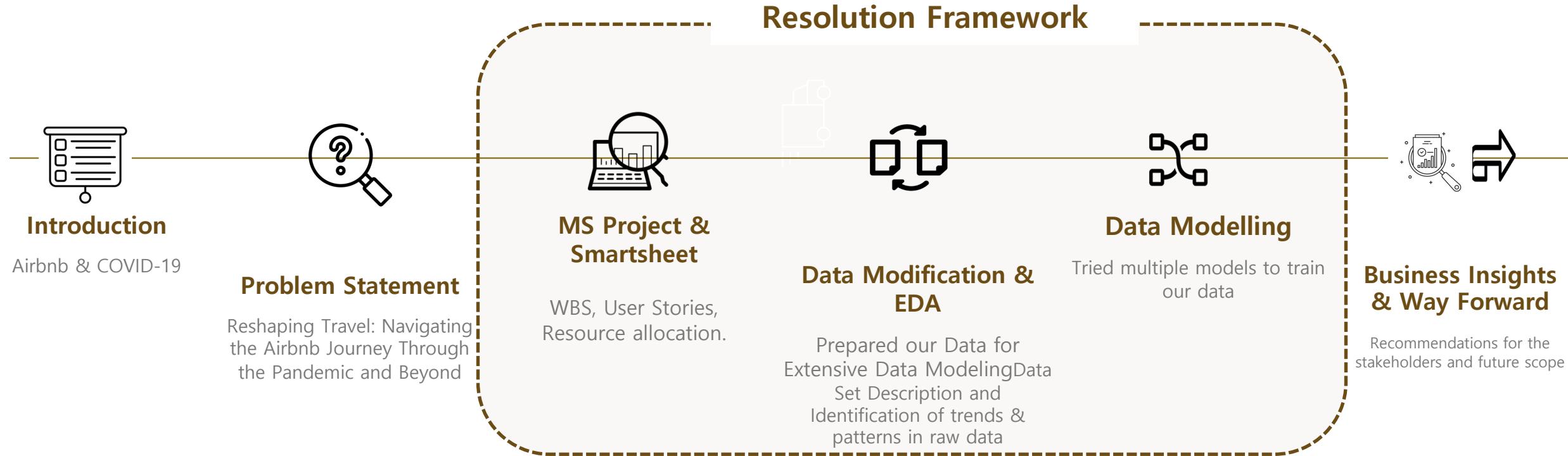




MGMT 585 - AIRBNB | IMPACT OF COVID ON AIRBNB BOOKINGS IN CHICAGO AND BOSTON

TEAM SCRUM SPRINTERS: AKANKSHA SINGH, CHINMAY GIDWANI, DEVARSHI SHARMA, HARSHRAJ JADEJA, KEERTANA MADAN

AGENDA



Project Objective

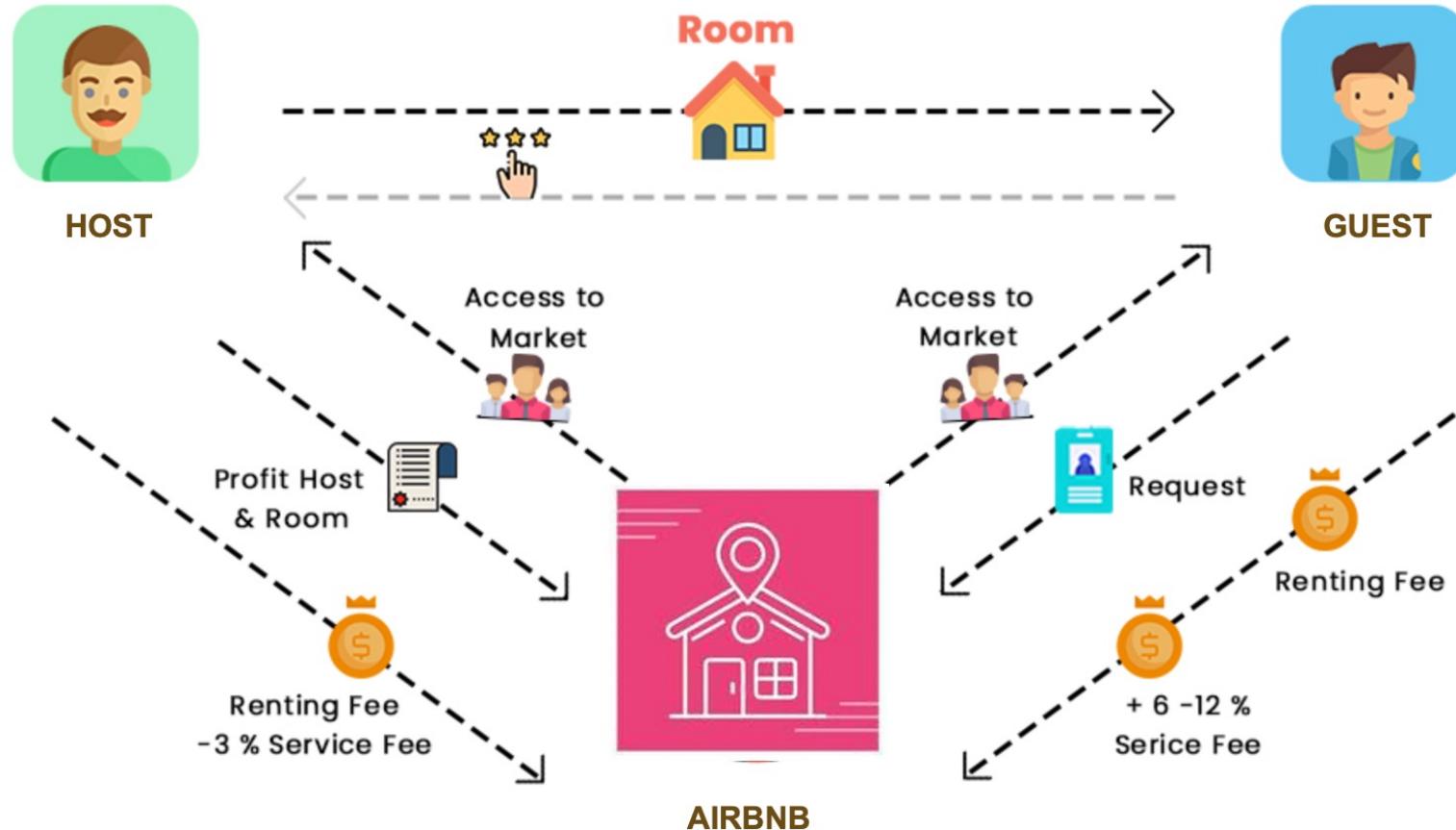
UTILIZE GIVEN DATA SETS ON AIRBNB TO SOLVE REAL TIME PROBLEMS USING ROBUST MODELING TECHNIQUES



INTRODUCTION

AIRBNB BUSINESS MODEL

Insight into Stakeholders within a Platform-Based Airbnb Business Operations.





PROBLEM IDENTIFICATION

PROBLEM STATEMENT

Reshaping Travel: Navigating the Airbnb Journey Through the Pandemic and Beyond



- The COVID-19 pandemic has exerted profound and multifaceted impacts on urban centers globally, with cities like Boston and Chicago experiencing unique challenges and responses due to their distinct socio-economic landscapes, population density, and governance structures.



- This study seeks to uncover and analyze the comprehensive impact of the COVID-19 pandemic on Boston and Chicago, focusing on key areas such as public health outcomes, economic disruptions and recovery efforts, changes in urban mobility and public transportation usage, the efficacy of social services and support systems in addressing the pandemic-induced needs, and the adaptive strategies implemented by local governments to manage the crisis.



- Additionally, the study aims to explore the shifts in community behavior, resilience, and innovation in response to pandemic-related restrictions and fears.



**Consumer Behavior Change? Rural vs Urban Demand? Hosts' Response Effective?
Airbnb's Strategic Shifts?**

Consumer Shifts: Unveiling a dramatic pivot in traveler preferences towards secluded, longer stays and work-friendly accommodations.

Geographic Redistribution: Highlighting a marked migration of bookings from bustling cities to tranquil, rural escapes.

Host Adaptability: Showcasing the resilience and innovation of Airbnb hosts in enhancing safety and flexibility for guests.

Business Evolution: Revealing Airbnb's strategic transformation to navigate market shifts and redefine its role in the future of travel.

Problems Identification



MS PROJECT & SMARTSHEET

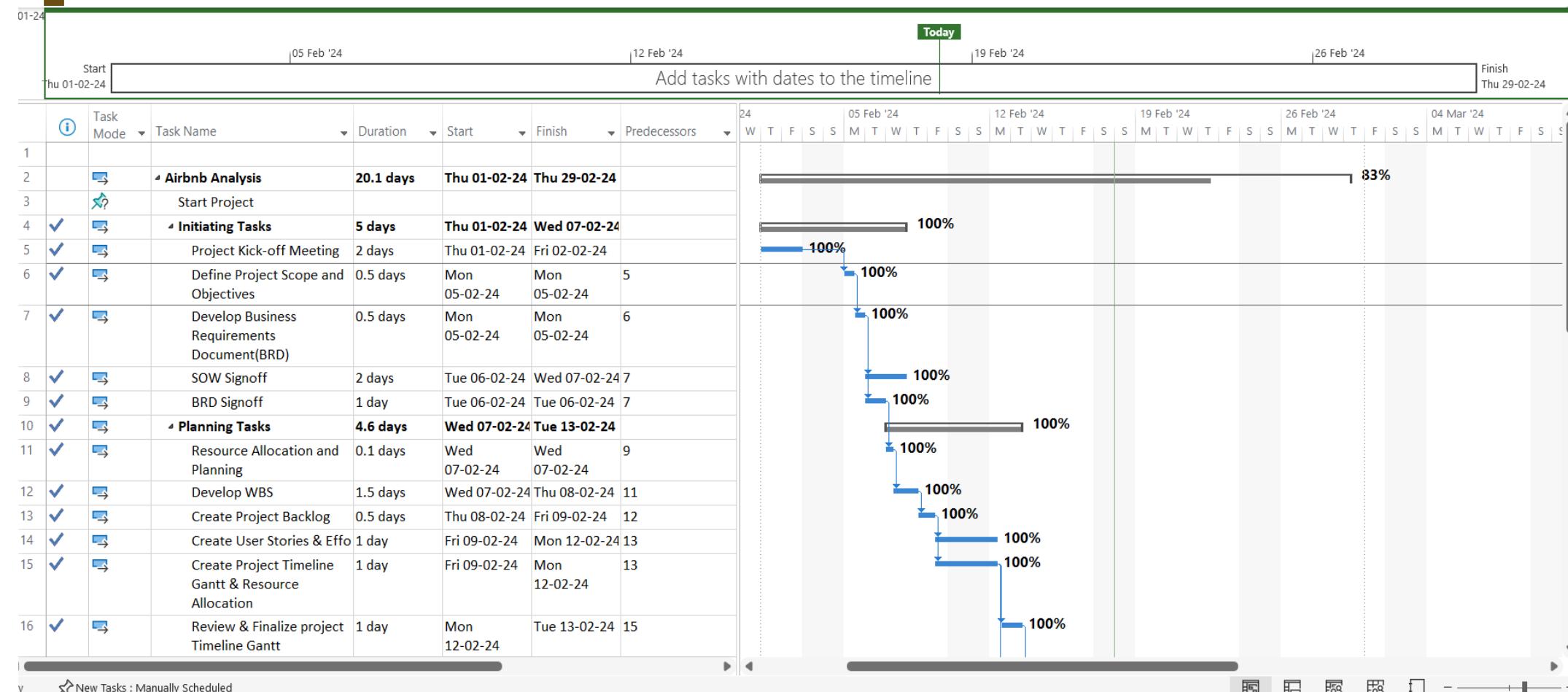
WBS



Implemented a comprehensive, multi-faceted strategy to pinpoint real-time challenges experienced by all stakeholders in a platform-based business model

	Task		Task(Cont'd)
1	Initiating Tasks	20	ADS completed
2	Project Kick-off Meeting	21	EDA (Covid Dataset)
3	Define Project Scope and Objectives	22	EDA (Airbnb Dataset)
4	Develop Business Requirements Document(BRD)	23	EDA & Data Analysis(ADS)
5	SOW Signoff	24	Preliminary Analysis Report & Visualizations
6	BRD Signoff	25	Hypothesis Testing
7	Planning Tasks	26	Feature Engineering
8	Resource Allocation and Planning	27	Feature Selection
9	Develop WBS	28	Preliminary Model Building
10	Create Project Backlog	29	Difference in Difference Modelling
11	Create User Stories & Effort Estimation	30	Airbnb & Covid Impact Analysis Report
12	Create Project Timeline Gantt & Resource Allocation	31	Monitoring & Controlling
13	Review & Finalize project Timeline Gantt	32	Review & Revise Model
14	Executing Tasks	33	Modelling Data Review Meeting
15	Data Collection on Airbnb Market Performance	34	Incorporate Feedback into model
16	Search for public data on covid	35	Closing Tasks
17	Finalize covid dataset	36	Create Project Documentation - Confluence
18	Data Cleaning & pre-processing	37	Presentation to Stakeholders
19	Preparation of ADS	38	Project Debrief

GANTT VIEW



USER STORIES & TASKS PLANNING



Edit

Title
Preliminary Analysis Report & Visualizations

Description
As a Data Analyst, I will synthesize insights from the exploratory data analyses of both Airbnb and COVID datasets into a preliminary report, complemented by visualizations, to provide an initial understanding of the pandemic's impact on Airbnb.

Acceptance Criteria
A preliminary analysis report is produced, incorporating clear and insightful visualizations (charts, graphs, heatmaps) that illustrate initial findings on the relationship between COVID-19 progression and Airbnb market changes. This report is to be reviewed and approved by the project team.

Add Attachment
Create Proof

Cancel OK

Row 13: Preliminary Analysis Report & Vis...

Keertana Madan
EDA Task was completed and the findings were reported in a Jupyter Notebook.
Just now
Reply

User Story Format

- Created Stories using As a, I want, So that
- Defined Acceptance Criteria for completion

Story Creation Workshops

- Conducted workshops
- Curated List of tasks for project
- Organised list into Stories & Tasks

app.smartsheet.com/sheets/841Hm6V8QG7j3F86mH43MhrFQV35xxc634jWXVC1?view=card&cardLevel=0&cardViewByColumnId=61500...

smartsheet

To Do (2) In Progress (2) Review/Test (3) Blocked (1) Done (16) + Add Lane

Edit

Title
Building Data Models

Description
As a Data Scientist, I will develop predictive models using the cleaned and integrated ADS to forecast Airbnb market performance under varying future COVID-19 scenarios, leveraging machine learning techniques.

Acceptance Criteria
At least one predictive model is built and demonstrates the ability to accurately forecast market trends based on different COVID-19 progression scenarios. Model performance metrics (e.g., RMSE, R-squared) meet the project's predefined benchmarks.

Priority

Add Attachment
Create Proof

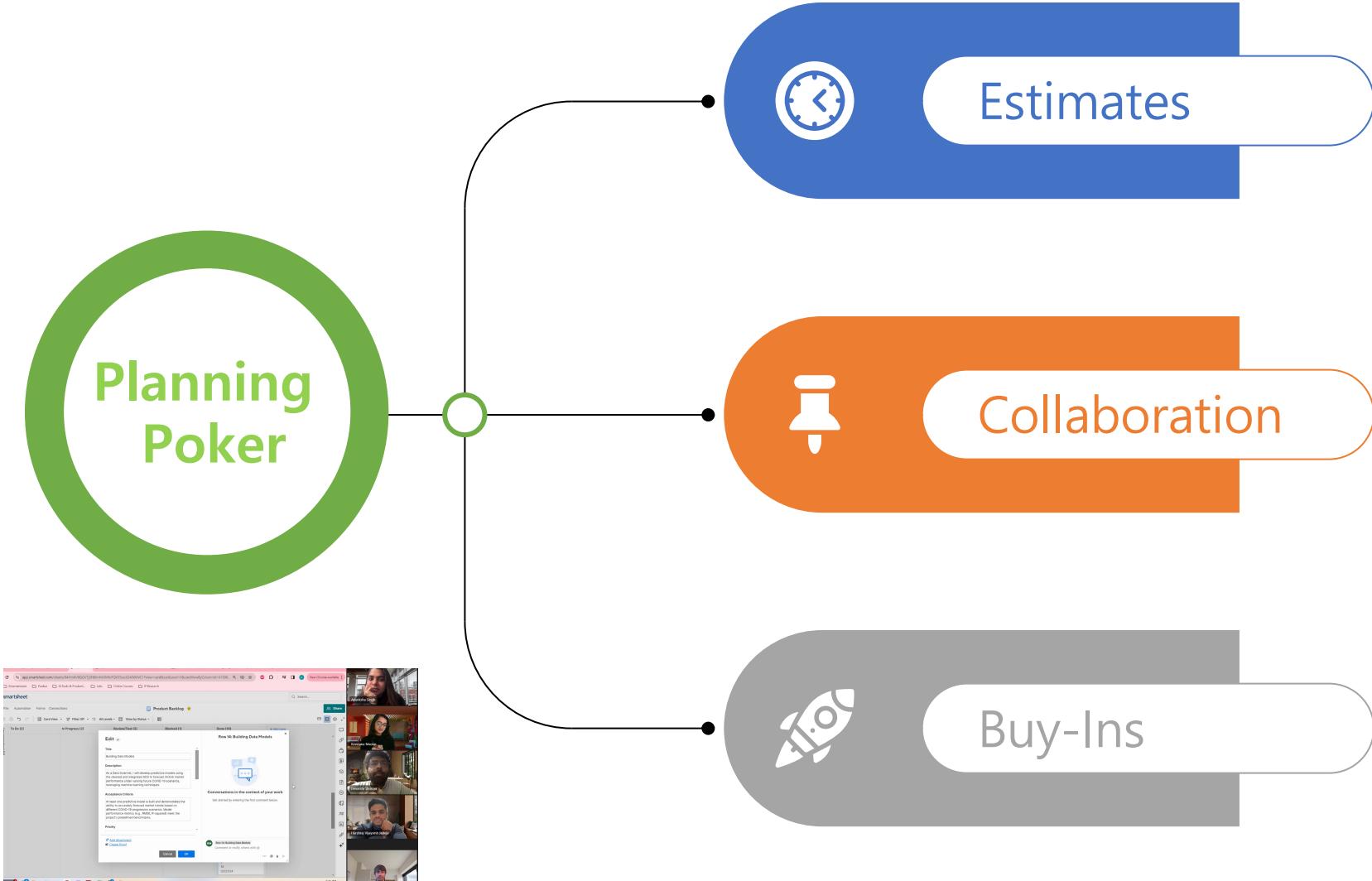
Cancel OK

Row 14: Building Data Models

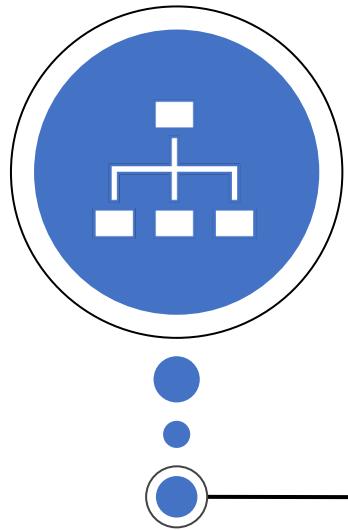
Conversations in the context of your work
Get started by entering the first comment below.

Alanksha Singh
Keertana Madan
Devarshi Sharma
Harshaj Vijaysinh Jadeja

EFFORT ESTIMATION

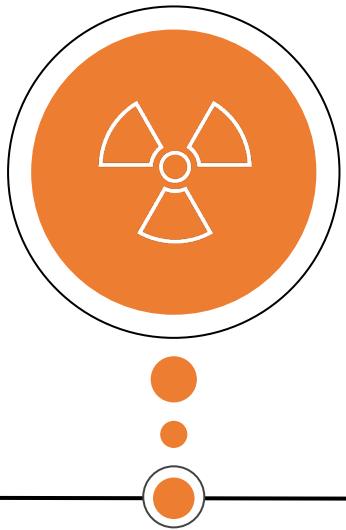


USER STORY PRIORITIZATION



Dependency Mapping

Dependency mapping influenced the prioritization to ensure a logical flow in project timeline



Risk Assessment

High risk tasks like "Hypothesis Testing" & "Feature Engineering" addressed early in timeline



Backlog

List of all stories and tasks ordered

PROJECT BACKLOG



7 User Stories, 17 Tasks

smartsheet

File Automation Forms Connections **Product Backlog** star

Search...

#	Title	Description	Acceptance Criteria	Priority	Story Points	Assigned To	Assign... %	Predeces...	Status	Type
4	- Preparation of Analytical Dataset(ADS)	As the Data Scientist Lead, I aim to consolidate historical data from various sources to create a central dataset.	An ADS is prepared and includes all relevant historical data.	1	5	Harshraj Jade	100%		Done	User Story
5	Data Collection on Airbnb Market Performance	As a Data Analyst, I want to gather historical market performance data from Airbnb's public API.	Relevant Airbnb market performance data is collected.	1	1	Keertana Mac	50%		Blocked	Task
6	Search for public data on covid	As a Data Scientist, I want to obtain public data on COVID-19 for analysis.	Comprehensive COVID public dataset is found.	1	1	Chinmay Gid	50%		Done	Task
7	Finalize covid dataset	As a Data Analyst, I want to select the most accurate COVID dataset for our analysis.	The COVID dataset is finalized.	1	1	Akanksha Sir	50%	6	Done	Task
8	Data Cleaning & pre-processing	As a Data Scientist, I want to clean and preprocess the selected COVID dataset.	Data is cleaned and pre-processed.	1	1	Keertana Mac	100%	7	Done	Task
9	ADS completed	As a Data Scientist, I want to merge the Airbnb and COVID datasets.	The ADS is fully merged, with COVID data included.	1	1	Chinmay Gid	100%	8	Done	Task
10	- EDA & Data Analysis(ADS)	As a Data Analyst, I want to perform a comprehensive analysis on the combined dataset.	EDA is completed on the ADS, visualizations are generated.	2	7	Akanksha Sir	100%		Done	User Story
11	EDA (Covid Dataset)	As a Data Analyst, I want to perform exploratory data analysis on the COVID dataset.	Exploratory Data Analysis (EDA) is performed on the COVID dataset.	1	2	Akanksha Sir	100%	8	Done	Task
12	EDA (Airbnb Dataset)	As a Data Analyst, I want to explore the Airbnb dataset.	EDA is performed on the Airbnb dataset.	1	2	Harshraj Jade	100%	8	Done	Task
13	Preliminary Analysis Report & Visuals	As a Data Analyst, I will synthesize insights from both datasets into a report.	A preliminary analysis report is generated.	1	3	Chinmay Gid	100%	11, 12	Done	Task
14	- Building Data Models	As a Data Scientist, I will develop predictive models based on the consolidated dataset.	At least one predictive model is built.	3	9				Done	User Story
15	Hypothesis Testing	As a Data Scientist, I want to conduct hypothesis testing on the data.	Hypotheses are tested, with results documented.	1	2	Akanksha Sir	20%	13	Done	Task
16	Feature Engineering	As a Data Scientist, I want to develop new features for the model.	New features are engineered and integrated.	1	1	Akanksha Sir	25%	15	Done	Task
17	Feature Selection	As a Data Analyst, I want to select the most relevant features for the model.	The most predictive features are selected.	1	1	Devarshi Sha	50%	15	Done	Task
18	Preliminary Model Building	As a Data Scientist, I want to build a preliminary model.	A preliminary model is built, and its performance is evaluated.	1	1	Chinmay Gid	100%	16, 17	Done	Task
19	Difference in Difference Modelling	As a Data Scientist, I want to apply Difference-in-Differences modeling.	A DID model is successfully implemented.	1	1	Harshraj Jade	100%	18	Done	Task
20	Review & revise Model	As a Data Scientist, I want to based on initial findings, review and refine the model.	The model is iteratively revised, and final results are presented.	1	3	Akanksha Sir	20%		Review/Test	Task
21	- Report Findings and Insights	As a Data Scientist, I want to consolidate the findings and present them in a report.	A detailed report is produced, containing key findings and recommendations.	4	4				Review/Test	User Story
22	Airbnb & Covid Impact Analysis Report	As a Project Manager, I want a comprehensive report on the impact of COVID on Airbnb.	A comprehensive analysis report is generated.	1	1	Keertana Mac	100%	20FS-1d	Done	Task
23	Modelling Data Review Meeting	As a Project Manager, I want to conduct a meeting to review the data modelling process.	A review meeting is conducted, and feedback is gathered.	1	1	Akanksha Sir	100%	22, 20	Done	Task
24	Incorporate Feedback into model	As a Data Analyst, I want to incorporate stakeholder feedback into the model.	Feedback is incorporated into the final model.	2	2	Harshraj Jade	100%	23	In Progress	Task
25	Create Project Documentation - Confluence	As a Data analyst, I want to develop a comprehensive project documentation.	Project documentation is thorough and up-to-date.	5	3	Chinmay Gid	50%	22	In Progress	User Story
26	Presentation to Stakeholders	As a Project Lead, I want to prepare and deliver a presentation to stakeholders.	A professional presentation is delivered.	5	0	Akanksha Sir	20%	22	In Progress	Task
27	Project Debrief	As a Project Lead, I want to conduct a debrief session with the team.	A project debrief is held, with lessons learned documented.	5	2	Akanksha Sir	20%	26	To Do	User Story

KANBAN BOARD



smartsheet

File Automation Forms Connections

Product Backlog ★

Card View Filter Off All Levels View by Status

To Do (1) In Progress (3) Review/Test (2) Blocked (1) Done (17)

Uncategorized (0)

To Do (1)	In Progress (3)	Review/Test (2)	Blocked (1)	Done (17)
Project Debrief 26 5 2 1d 02/29/24	Report Findings and Insights Incorporate Feedback into model 23 2 2d 02/27/24	Building Data Models Review & revise Model 3 4d 02/21/24	Preparation of Analytical Dataset(ADS) Data Collection on Airbnb Market Performance 1 2d 02/12/24	Preparation of Analytical Dataset(ADS) Market Performance 1 5 5d 02/12/24
Create Project Documentation - Confluence 22 5 3 2d 02/27/24	Report Findings and Insights 4 4 3d 02/26/24	+ + + +	Preparation of Analytical Dataset(ADS) Search for public data on covid 1 1d 02/12/24	Report Findings and Insights Modelling Data Review Meeting 22, 20 1 0 02/26/24
Presentation to Stakeholders 22 5 0 2d 02/27/24	+ + + +	+ + + +	Preparation of Analytical Dataset(ADS) Finalize covid dataset 6 1 1d 02/13/24	Preparation of Analytical Dataset(ADS) Data Cleaning & pre-processing 1 1d 02/13/24

3 In Progress

17 Completed

1 Blocked

1 To Do

DATA SET INFORMATION - AIRBNB

Implemented a comprehensive, multi-faceted strategy to pinpoint real-time challenges experienced by all stakeholders in a platform-based business model



No. of Rows & Columns

121K Rows & 111 Columns



Average Booked Days

26 Days Booked



Time Period

16 periods



Average Occupancy Rate

19% Occupancy



Total No. of Hosts

13.6 K Hosts



Average Revenue

\$3,301



Super Hosts Status

~4,600 Hosts



Average Rating Overall

94%

DATA SET INFORMATION - COVID

Implemented a comprehensive, multi-faceted strategy to pinpoint real-time challenges experienced by all stakeholders in a platform-based business model



Chicago
2,697,000 population



No. of Cases
320,000



Time Period
2 periods



No of Deaths
7,150



Boston
654,776 population



No. of Cases
77,500



Time Period
2 periods



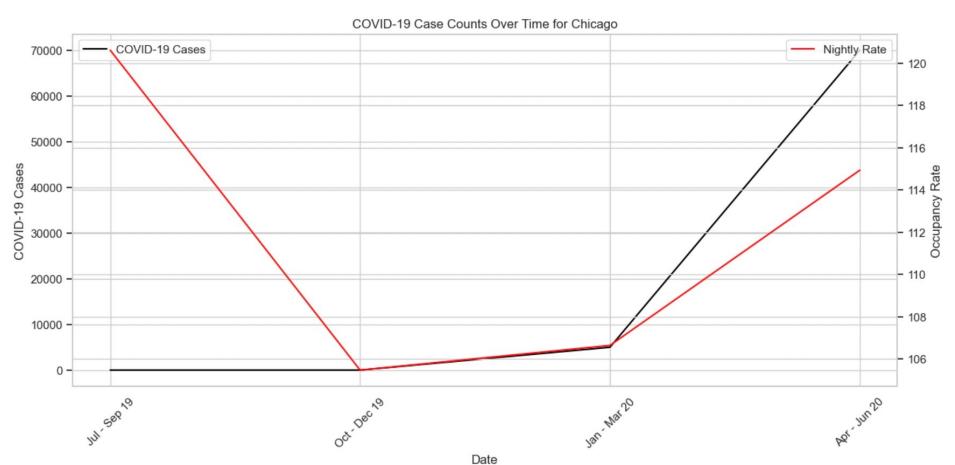
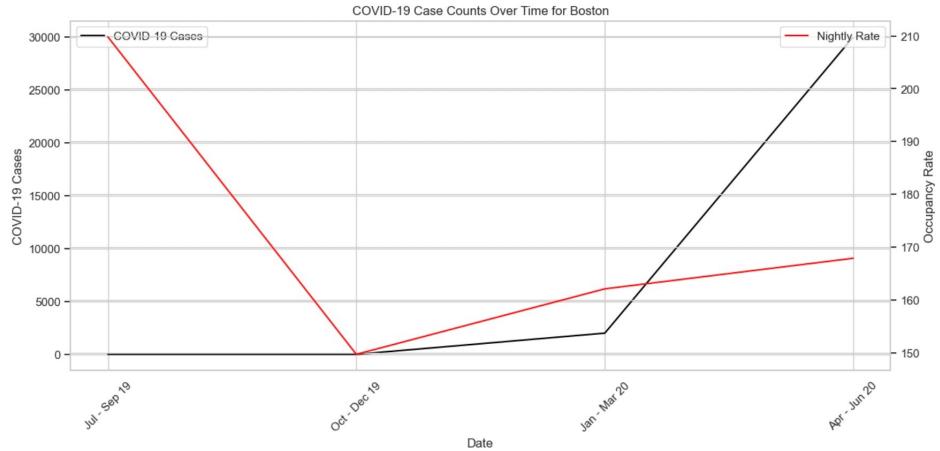
No of Deaths
2,170



EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis indicating the COVID-19 case count over time.



From the first graph for Chicago:

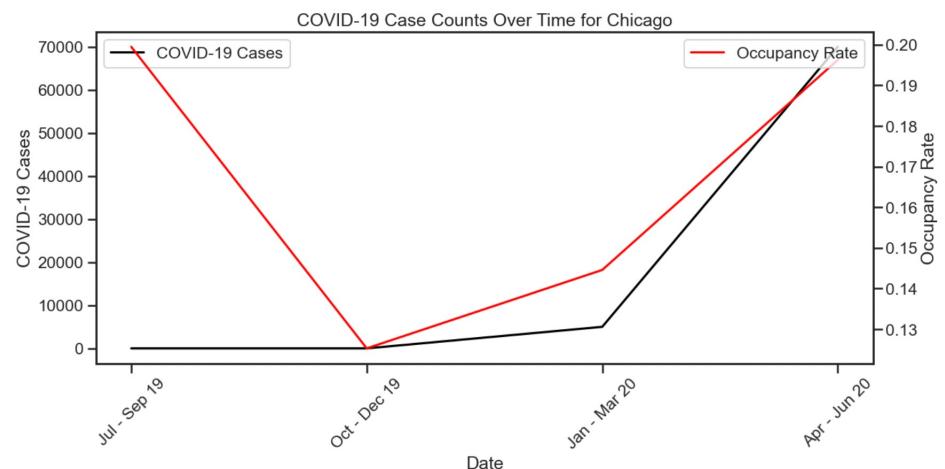
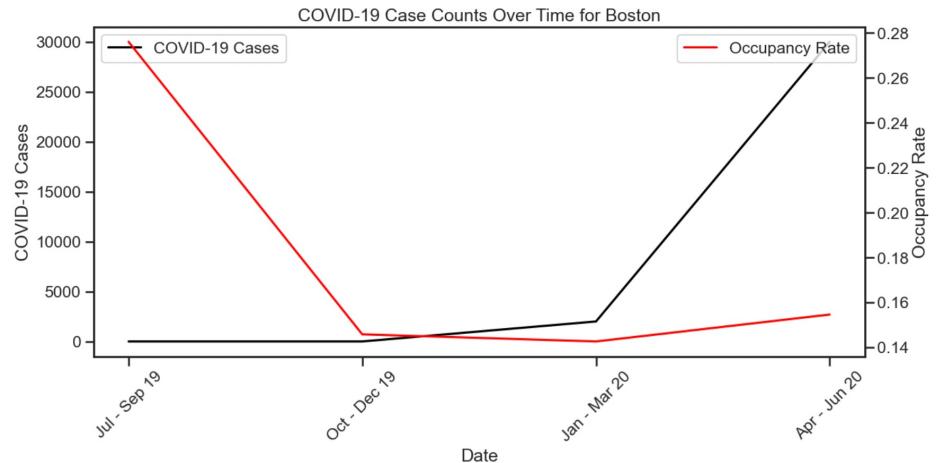
The graph shows a sharp increase in nightly rates that coincides with a decrease in COVID-19 cases, indicating a potential correlation where the decrease in cases might contribute to a higher willingness to pay more for accommodations.

From the second graph for Boston, we can infer:

The trend is similar, with the nightly rates increasing as COVID-19 cases decline. This could reflect a recovery in the travel and accommodations sector, with higher rates suggesting a rebound in demand as the perceived risk associated with traveling decreases.

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis indicating the COVID-19 case count over time.



From the first graph for Boston:

Correlation between Cases and Occupancy: There seems to be an inverse relationship between the rise of COVID-19 cases and the occupancy rate, suggesting that as cases increased, the occupancy rates decreased.

Recovery Signs: The occupancy rate begins to recover around April - June 2020, which could indicate a possible adaptation to the pandemic conditions or the implementation of safety measures that restored traveler confidence.

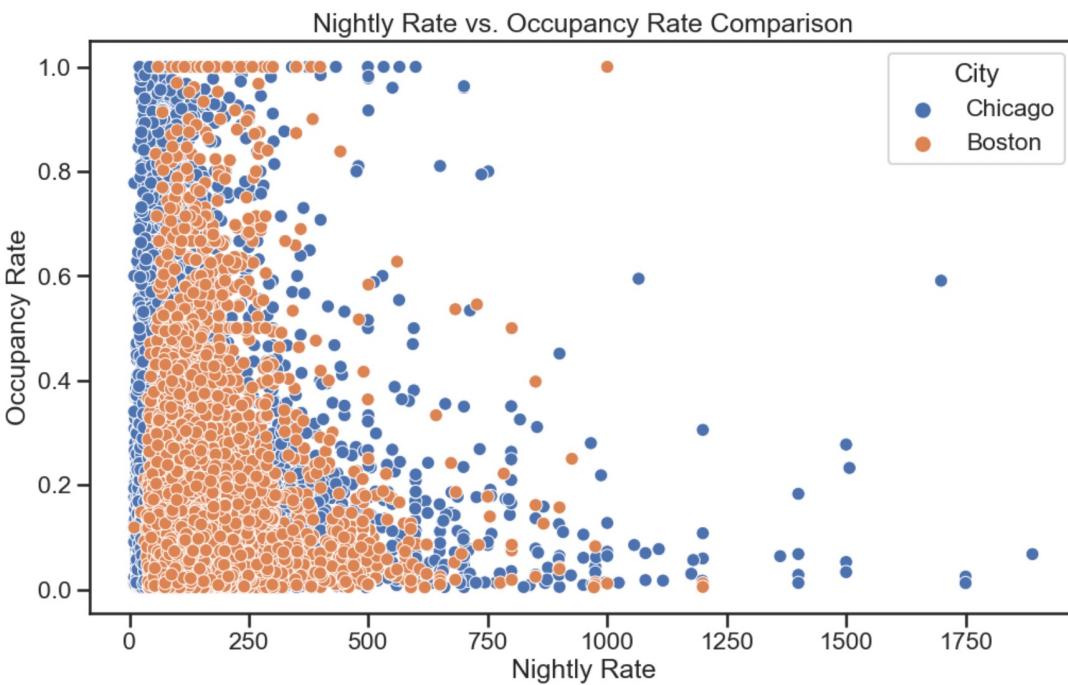
From the second graph for Boston, we can infer:

Greater Impact on Occupancy: Chicago's occupancy rates also show a decline as COVID-19 cases rise, but the drop appears more significant compared to Boston, which might reflect stricter lockdown measures or a higher impact of the pandemic in this city.

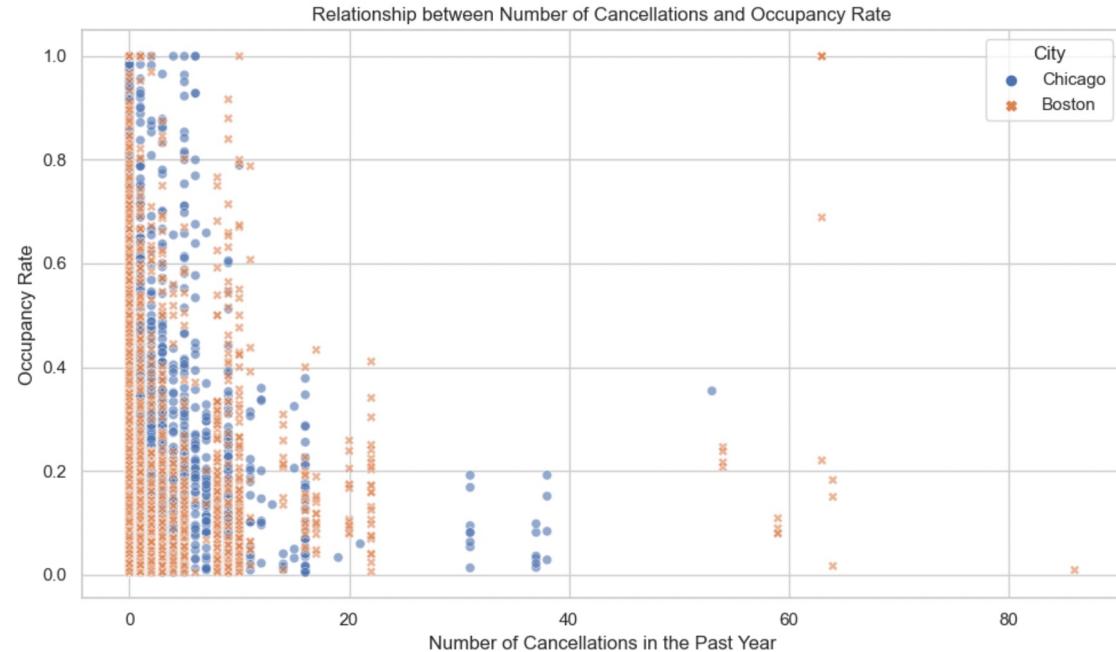
Rapid Rebound: Like Boston, there is a rebound in occupancy rates as cases start to decline. However, the recovery in Chicago seems to be more pronounced and rapid, potentially due to various factors like the lifting of restrictions or seasonality effects.

EXPLORATORY DATA ANALYSIS

Nightly Rate vs Occupancy Rate & Number of Cancellations vs Occupancy Rate



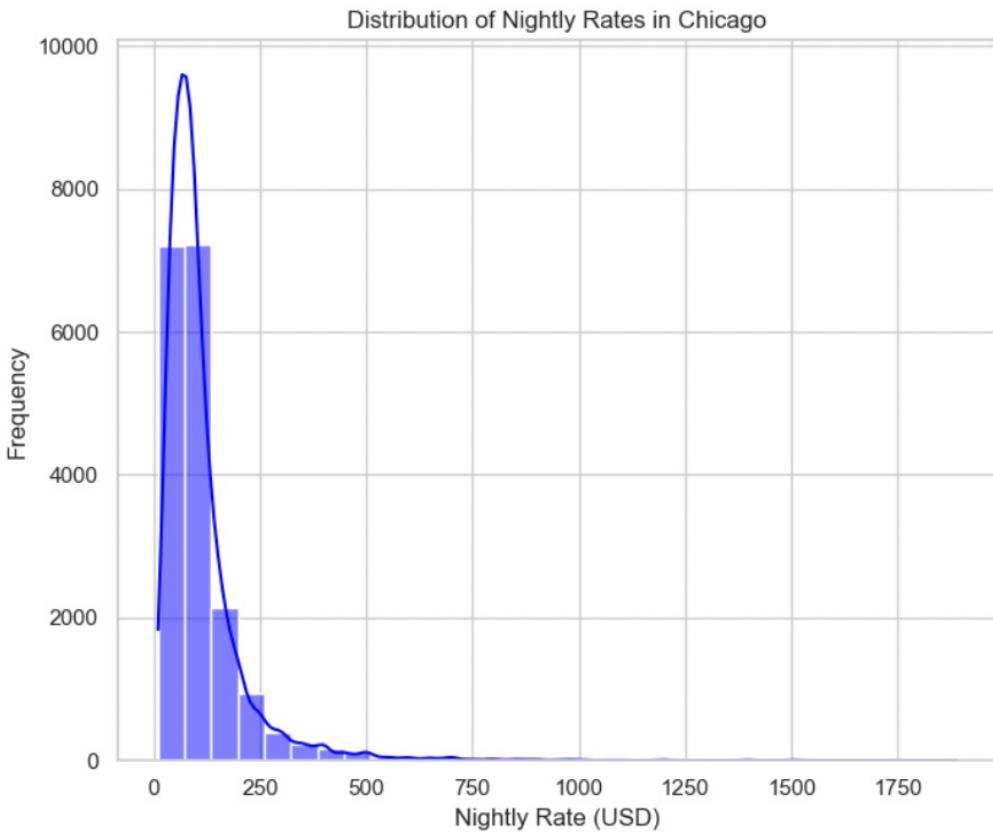
The scatter plot suggests that lower-priced Airbnb listings generally achieve higher occupancy rates, with Chicago showing a denser concentration of high-occupancy, lower-cost options compared to Boston.



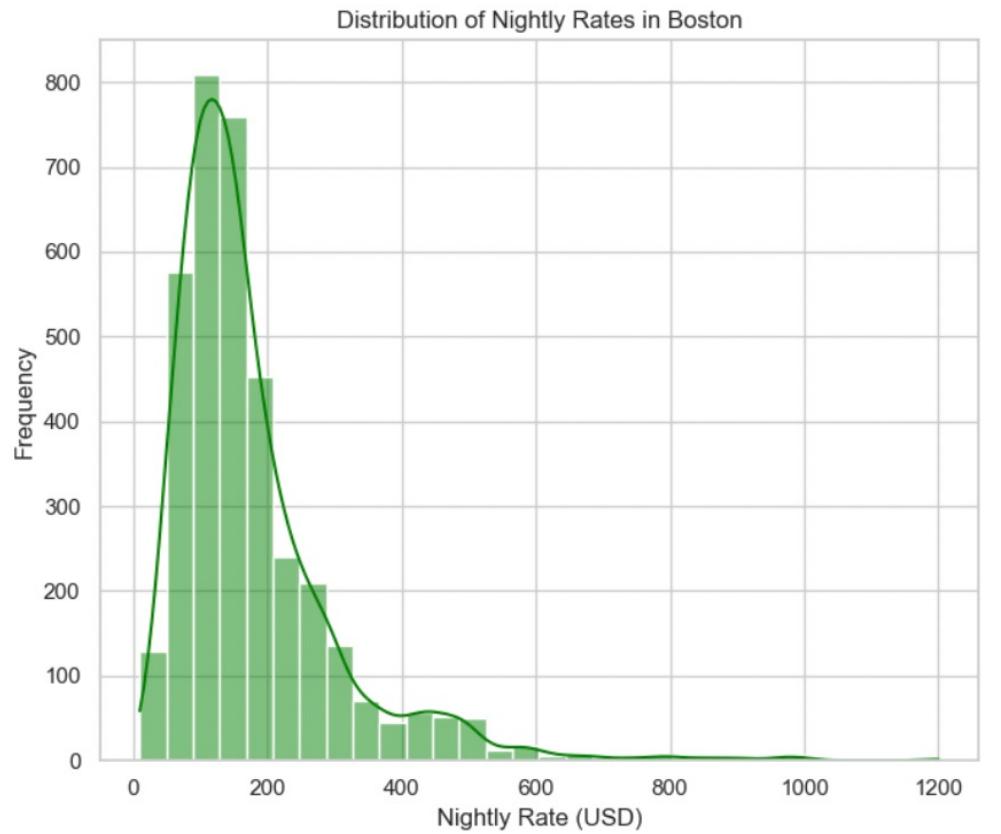
The scatter plot suggests a common trend in both Chicago and Boston, where Airbnb listings with fewer cancellations tend to have higher occupancy rates.

EXPLORATORY DATA ANALYSIS

Distribution of Nightly Rates



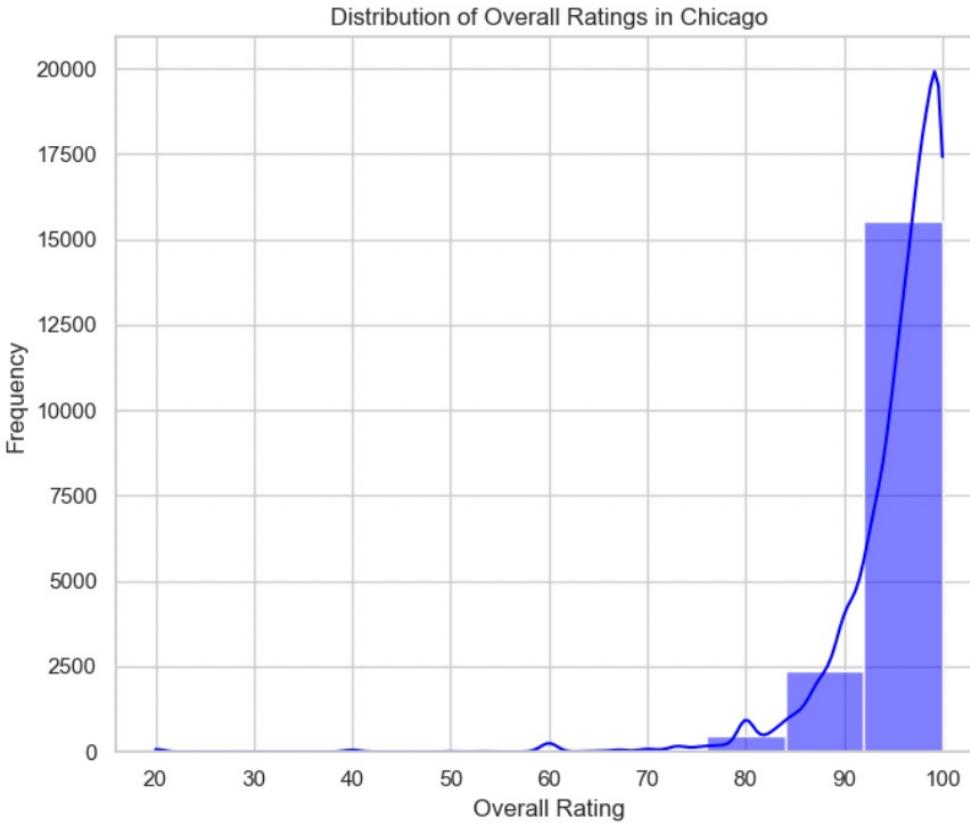
The histogram for Chicago indicates a high frequency of Airbnb listings concentrated at lower nightly rates, with a steep drop-off as prices increase



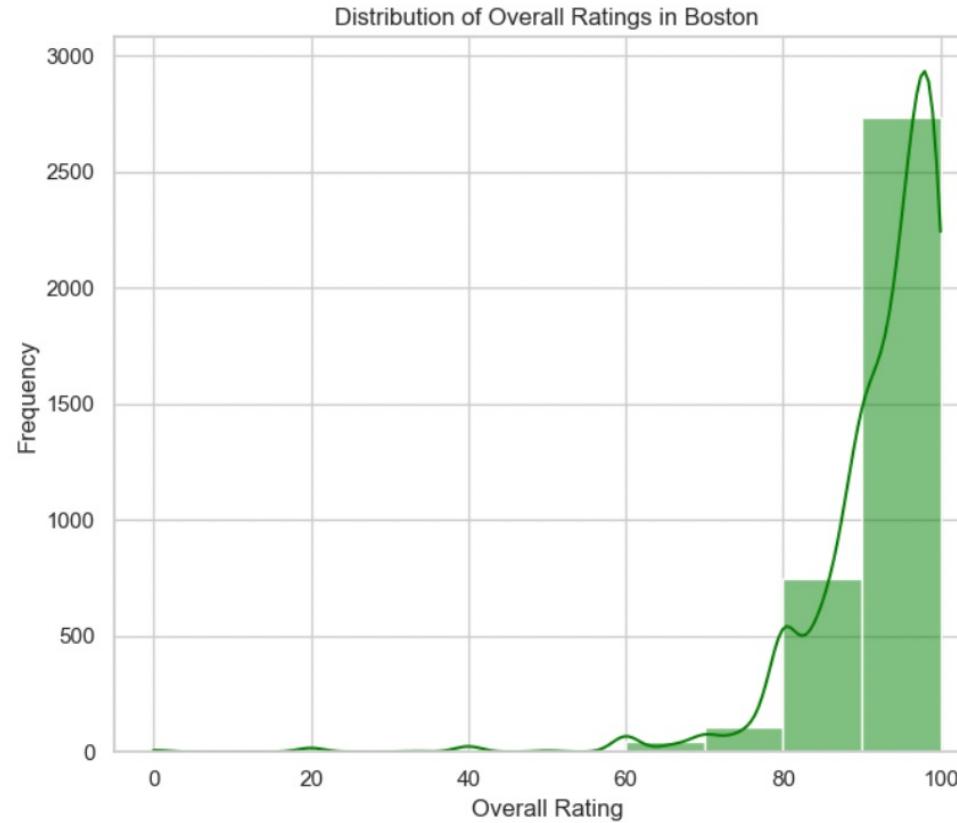
Boston's distribution shows a broader range of nightly rates with a more gradual decline in frequency as rates rise.

EXPLORATORY DATA ANALYSIS

Distribution of Overall Ratings



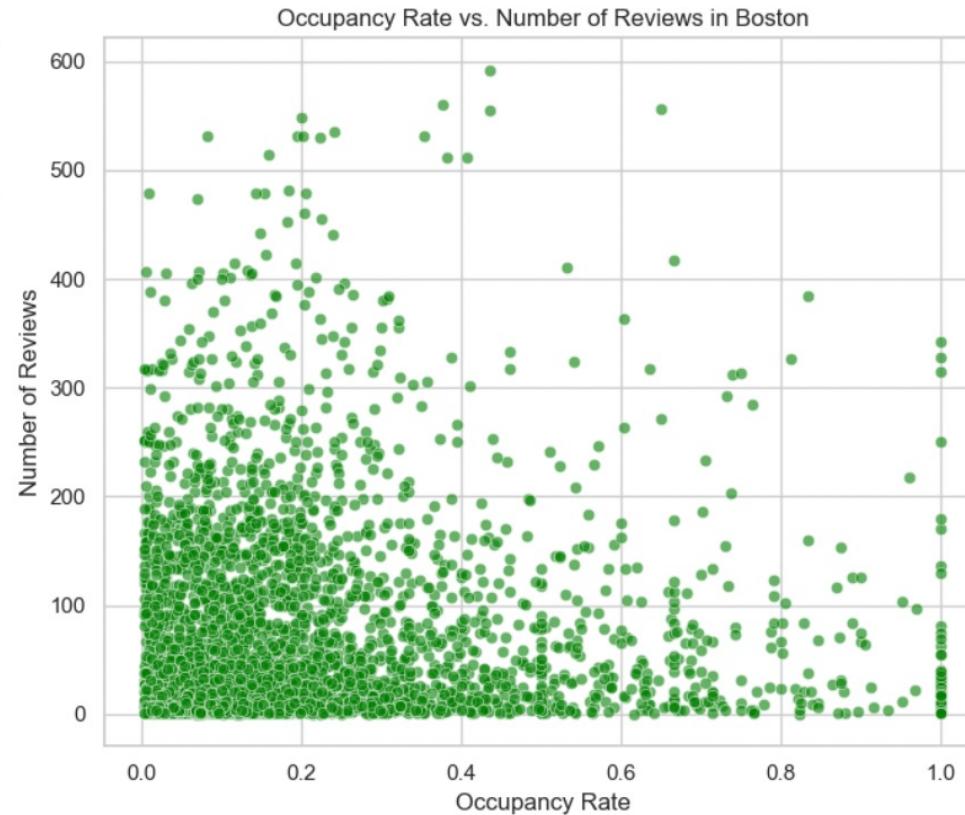
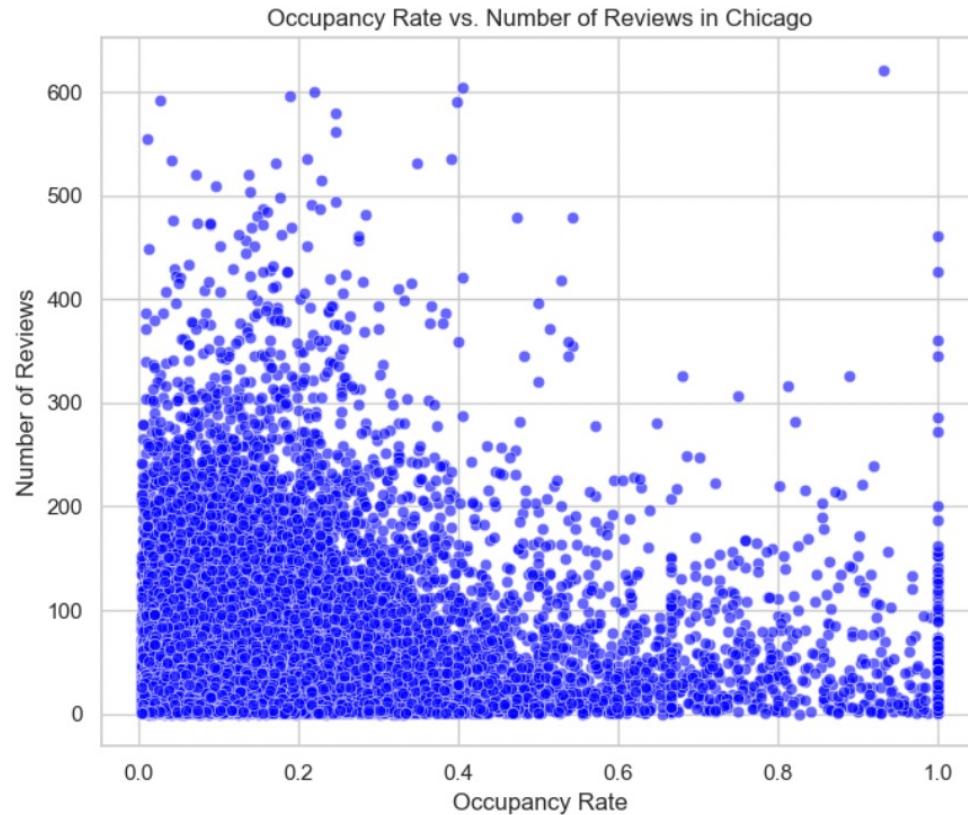
The histogram for Chicago shows a predominant concentration of Airbnb listings with high overall ratings, skewing towards the upper end



Boston's distribution also suggests a high frequency of listings with top ratings, with a slightly more even spread across the rating spectrum.

EXPLORATORY DATA ANALYSIS

Occupancy Rate vs Number of Reviews



In the scatter plots, Chicago's Airbnb listings show a dense cluster at lower occupancy rates regardless of the number of reviews

Boston's plot displays a more even spread across occupancy rates with a similar pattern of reviews.

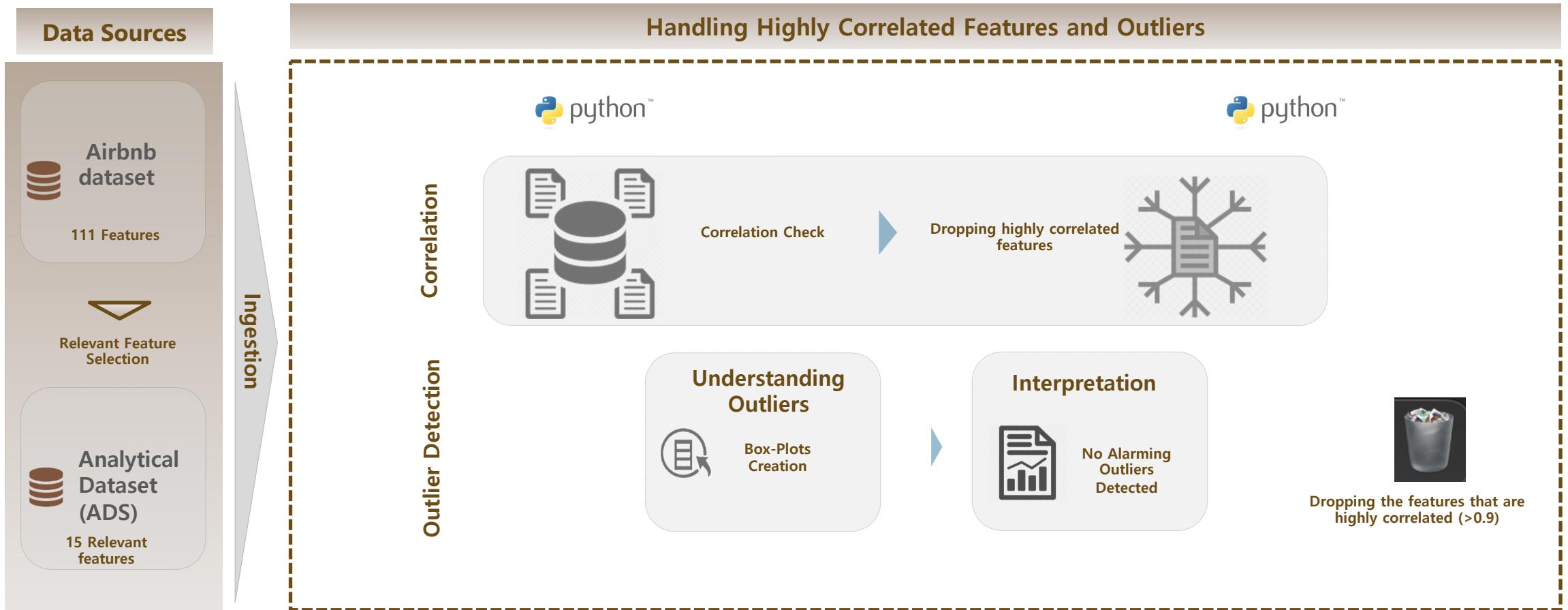


DATA MODIFICATION



DATA PREPROCESSING

Layer 1 – Feature Selection & Data Ingestion | Removal of irrelevant features, to increase reliability



DATA PREPROCESSING

Layer 2 – Handling Missing Values | A significant proportion of the dataset contains nulls, hence, making this layer significant



Handling Missing Values

Feature	Null count	Null percent
Property Type	3	0%
Minimum Stay	3	0%
Number of Photos	8	0%
Max Guests	11	0%
Bedrooms	17	0%
Number of Reviews	27	0%
Bathrooms	88	0%
Neighborhood	3003	2%
numReserv_pastYear	7398	4%
hostResponseAverage_pastYear	12074	7%
available_days	23449	14%
available_days_aveListedPrice	23449	14%
Rating Overall	24372	15%
Cleaning Fee (USD)	32537	19%
numCancel_pastYear	34466	21%
numReviews_pastYear	34466	21%
prop_5_StarReviews_pastYear	35856	21%
rating_ave_pastYear	35856	21%
booked_days	55982	33%
occupancy_rate	55982	33%
revenue	55982	33%



Dropping Null records with low proportion



Dropping Null Records where Prediction/Imputation can potentially bias the dataset

Nulls remaining

	hostResponseAverage_pastYear	977
	Cleaning Fee (USD)	11730

Gradient Boosting | Null Imputation

```
# List of columns to impute
columns_to_impute = ['hostResponseAverage_pastYear', 'Cleaning Fee (USD)']

# Initialize IterativeImputer with GradientBoostingRegressor
imputer = IterativeImputer(estimator=GradientBoostingRegressor(), max_iter=10, random_state=0)

# Fit and transform the data
data[columns_to_impute] = imputer.fit_transform(data[columns_to_impute])
```

Gradient boosting is preferred over mean imputation because –

- Iterative learning process
- Can model complex relationships in the data
- Captures non-linear relationships between features
- Generally robust to outliers

Dropping Null Records

Imputing nulls using Gradient Boosting

DATA PREPROCESSING

Layer 3 – Multicollinearity Check | Removal of multicollinear features to increase the model accuracy and to mitigate the issue with correlated predictor variables



Handling Multicollinearity

	Variable	VIF
0	rating_ave_pastYear	112.704997
15	prop_5_StarReviews_pastYear	65.490117
4	available_days	17.319020
17	superhost_period_all	11.073625
8	Max Guests	10.148840
14	Bedrooms	9.901874
12	Nightly Rate	7.101657
13	occupancy_rate	7.007180
3	hostResponseAverage_pastYear	6.890141
7	booked_days	6.699808
16	Bathrooms	6.280258
9	Cleaning Fee (USD)	5.831461
6	available_days_aveListedPrice	4.884304
11	Number of Photos	3.987684
1	numReviews_pastYear	2.509902
5	numReserv_pastYear	2.415193
18	Superhost	2.399962
2	numCancel_pastYear	1.448262
10	Minimum Stay	1.180993

Removing
rating_ave_pastYear feature to
account
multicollinearity

	Variable	VIF
14	prop_5_StarReviews_pastYear	19.851328
3	available_days	12.840673
16	superhost_period_all	10.154257
7	Max Guests	10.139134
13	Bedrooms	9.899102
11	Nightly Rate	7.064871
2	hostResponseAverage_pastYear	6.555453
6	booked_days	6.431679
15	Bathrooms	6.191153
12	occupancy_rate	5.889711
8	Cleaning Fee (USD)	5.831257
5	available_days_aveListedPrice	4.855148
10	Number of Photos	3.986968
0	numReviews_pastYear	2.502118
4	numReserv_pastYear	2.398725
17	Superhost	2.254689
1	numCancel_pastYear	1.448202
9	Minimum Stay	1.175297



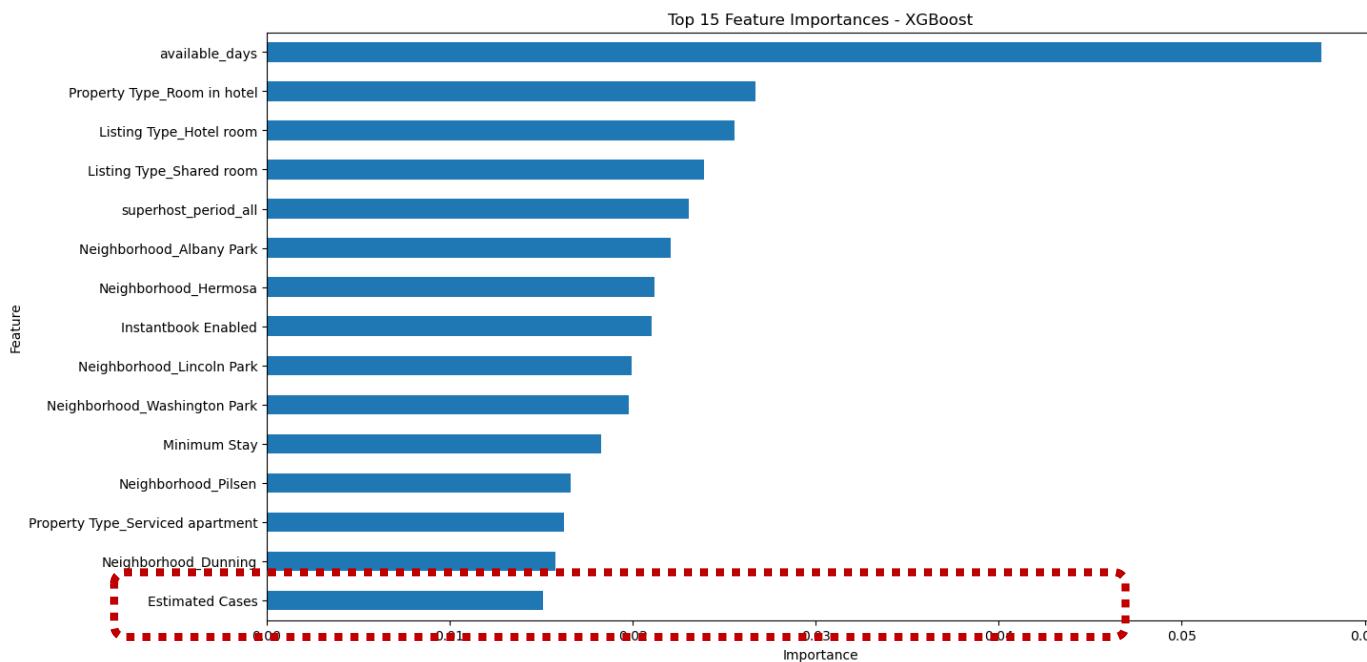
DATA MODELING

DATA MODELLING | TO IDENTIFY FEATURE IMPORTANCE



Utilizing XG Boost for data modeling revealed key feature importance that drive the performance of Airbnb listings (target variable being Occupancy Rate)

Feature Importance



Key Takeaway

- The bar chart represents the feature importance scores derived from an **XGBoost** machine learning model
- Primary Predictor** - Availability: 'available_days' emerges as the top predictor, indicating its critical role in determining the model's outcomes.
- Health Data Relevance:** The 'Estimated Cases' signifies that there is very little relevance of COVID-19, to the occupancy rate.

*RMSE: 0.133
*MAE: 0.088

DATA MODELING | DIFFERENCE IN DIFFERENCES ANALYSIS



The model summarizes a Difference in Differences analysis showing the impact of COVID-19 on Airbnb, with Boston experiencing a decrease in occupancy rates and revenue, while Chicago saw an increase in occupancy but a decrease in available days and booked days.

Treatment Group – The Airbnb listings after Covid cases started rising.

Control Group – The Airbnb listings before there were any Covid cases.

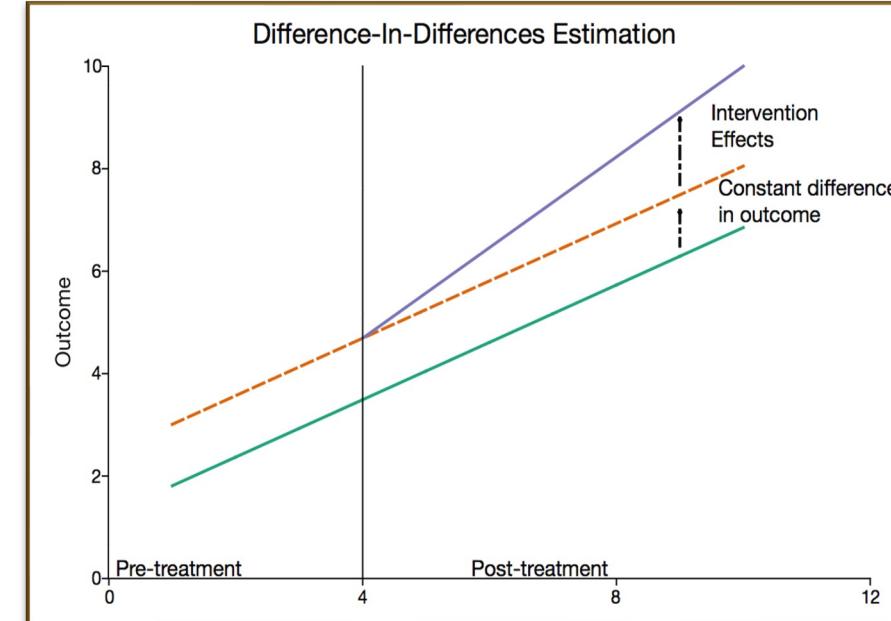
Treatment timeline:

Pre-Treatment – Jul '19 to Dec '19

Post-Treatment – Jan '20 to Jun '20

Average Occupancy Rate

Metric	Boston DID	Chicago DID
Booked Days	-4.183981	-3.390298
Occupancy Rate	-0.062353	0.008024



The formula for calculating the Difference in Differences (DiD) is:

$$\text{DiD} = (\text{Treatment Group Post-Treatment} - \text{Treatment Group Pre-Treatment}) - (\text{Control Group Post-Treatment} - \text{Control Group Pre-Treatment})$$

Implemented a Difference in Differences approach to calculate the treatment effect of covid on Airbnb



BUSINESS INSIGHTS & WAY FORWARD

BUSINESS INSIGHTS & RECOMMENDATIONS

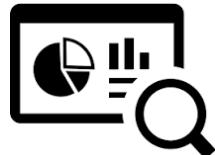
Comprehensive Business Insights to deal with real-time challenges experienced by all stakeholders in a platform-based business model



- Data-Driven Approach: The project emphasizes the use of robust data modeling to understand real-time challenges and Airbnb's response to the pandemic.
- Adaptive Methodologies: Agile practices like scrums and sprints and project management tools like MS Project and Smartsheet have been crucial for timely delivery and responsiveness.
- Risk Management: Early identification and management of high-risk tasks such as hypothesis testing and feature engineering have been pivotal.
- Collaborative Estimation: Team collaboration in effort estimation and consensus-building on project timelines has supported project cohesion.
- Quality Control: Addressing multicollinearity and outliers has ensured the integrity and reliability of data modeling.
- Change Management: The project includes strategies for hosts and recommendations for stakeholders, highlighting the importance of adaptability and continuous improvement.

FUTURE SCOPE

Potential areas of improvement and expansion in the future phases



Analysis



Project Management

Expand Geographic Analysis: Extend the study to include more cities or countries to compare Airbnb's impact and recovery patterns globally.

User Behavior Analysis: Deep dive into changing consumer preferences, focusing on the shift towards remote work, longer stays, and preferences for rural over urban settings.

Agile Methodology Enhancement: Refine agile methodologies used, incorporating more frequent feedback loops with stakeholders to adapt quickly to changes in the project's scope or market dynamics.

Longitudinal Study: Conduct a longitudinal study to track changes over a longer period, capturing post-pandemic recovery phases and new trends emerging in the travel industry.

Competitive Analysis: Compare Airbnb's performance and strategies with other accommodation and travel platforms to identify unique challenges and opportunities.

Technology Integration: Explore the use of emerging technologies (e.g., AI, blockchain) to enhance project management processes, such as automating routine



THANK YOU!