Analytics Startup Plan

Name: Hosain Ahmed
Student ID: 301209637

Synopsis: ***This document provides a high-level walkthrough of the activities required to guide completion of the analysis.***

Project	Seasonality Analysis of Airbnb in NYC
Requestor	Airbnb
Date of Request	07/04/2022
Target Quarter for Delivery	08/21/2022
Epic Link(s)	Not Available
Business Impact	Will try to figure predict the demand of Airbnb at different time of the year and factors responsible for the fluctuations.

1.0 Business Opportunity Brief

The aim of this study is to identify the important features in determining the availability of Airbnb properties in New York City. After going through the annual report of Airbnb, I found that the company is failing to generate enough revenue to be profitable. Moreover, the most significant expense of Airbnb is the cost of revenue. Therefore, Airbnb requires a model which will give the company more control over the knowledge of the more demanding areas where the properties will be booked more frequently and incur less cost to generate the revenue.

The specific ask:

- Which variables are responsible for availability of the properties of Airbnb throughout the year in New York?

1.1 Supporting Insights

Value Chain of Airbnb: Airbnb is a global online travel agency who generate revenue by renting out various privately owned properties and collecting service fee in return from the guests and hosts. Moreover, Airbnb generate more than 50 percent of its revenue from North American region. Airbnb is still facing the challenge of developing its service offerings as there is still scopes to educate both hosts and guests of Airbnb about the benefits of its services.

Competitors: Top competitors of Airbnb in terms of global revenue generations are Booking, Expedia, Trip.com, TripAdvisor and Trivago. However, among these competitors only Airbnb was able to show a positive revenue growth during the period from 2019 to 2021. (Statista, Topic: Airbnb)

1.2 Project Gains

Benefits of this research: This research will give Airbnb a model that will help the company in following business decisions:

- In which places and during which period of the year the demand is higher.
- It will help Airbnb design its product offerings and promotion at different periods of the year.

Implications of doing nothing:

- Airbnb will have excess capacity due to poor knowledge regarding seasonal demand.

2.0 Analytics Objective

Objective: Building a model to predict the seasonality of demand for Airbnb in New York City.

The specific Objectives:

- What is the impact of price over property availability?
- What property types are more popular and does this demand varies based on neighborhood?
- Does travel length or seasonality has any impact of property availability?

2.1 Other related questions and Assumptions:

Compared to its competitors, Airbnb is relatively new in the industry and it with a unique service offering. However, with the uniqueness of the service also calls for continuous product development, educating customers and understanding the demand of the customers. For instance, Airbnb has spent a significant portion in product development in last five years. Therefore, I believe this model will facilitate Airbnb with a tool to predict the demand of its service in a better way to take better business decisions in future. (Statista, Topic: Airbnb)

2.2 Success measures/metrics

The income statement of Airbnb shows that, between the years 2019-2021 the company has made a significant amount of loss. One of the major reasons for loss is the increasing cost of revenue. On the other hand, the dataset on hand shows that most of the properties of Airbnb remain available throughout the year in New York city, which is a tourist hotspot. Hence, we can infer that Airbnb has excess capacity in some neighborhoods which can be utilized properly to reduce the cost of revenue. This research will try to find the best model for features selection. Important variables will be identified to develop the predictive model for the rate of property availability. Airbnb can utilize the model to reduce its excess capacity, cost of revenue, and for product development.

2.3 Methodology and Approach

Type of Analysis:

After conducting the exploratory data analysis, I am planning to use following methods to find out the factors that are more relative in determining the property availability.

- Decision Tree
- Logistic Regression
- Neural Network

Methodology:

I will conduct exploratory data analysis by using bar charts to look for missing values, outliers and skewness in the data. After that, I will partition the data into train and validation. Then I will build a decision tree based on the sample, and observe which

variables are the most important in determining property availability. I will continue my analysis using logistic regression and neural networks to check whether the results support my prior findings. The reason for using various models to check if all models identify the same variables as responsible for the fluctuation of property availability.

Output:

With the findings I am hoping to identify the major factors for property availability which will help Airbnb to plan its future business decisions.

3.0 Population, Variable Selection, considerations

1. **Audience/population selection:** Guests and Hosts of Airbnb.
2. **Exclusions:** the following variables are excluded from the model:

EXCLUSION	VARIABLE	DEFINITION
Irrelevant variables	id	Property ID
	Name	Property Tag Name
	host_id	Property host ID
	last_review	When the last time host recieved a review
Variables with more skewness value of 7.56774264	calculated_host_listings_count	Number of property a single host is offering

The UNIVARIATE Procedure
Variable: calculated_host_listings_count

Moments			
N	49080	Sum Weights	49080
Mean	7.66105005	Sum Observations	376004.336
Std Deviation	34.7920617	Variance	1210.48756
Skewness	7.56774264	Kurtosis	61.1619255
Uncorrected SS	62290106.9	Corrected SS	59409518.8
Coeff Variation	454.142206	Std Error Mean	0.15704636

3. **Data Sources:**

- The “Airbnb” dataset was retrieved and downloaded from Kaggle. The dataset contains 48,895 entries with 16 columns including id, name, host_id, host_name, neighbourhood_group, neighbourhood, latitude, longitude, room_type, price,

minimum_nights,number_of_reviews,last_review,reviews_per_month,calculated_host_listings_count, availability_365

- Kaggle: https://www.kaggle.com/datasets/sakhawat18/asteroid-dataset

4. **Audience Level:** Guests and Hosts of Airbnb.
5. **Variable Selection:** the variables selected to run the model are the following:

VARIABLE	DEFINITION
neighbourhood_group	Larger neighborhood groups
neighbourhood	Specific áreas under the neighborhood groups
latitude	Latitude of the property
longitude	Longitude of the property
room_type	Types of rooms offered
price	Price per night stay in the property.
minimum_nights	Number of night stayed in the property
number_of_reviews	Total number of reviews each host received
reviews_per_month	Reviews received per month by each host
availability_365	The property was available for booking for how many days during the whole year

6. **Assumptions and data limitations:**

   Assumptions: Guests origin is anonymous and hosts are from New York City.

   Limitations: Missing values and skewness in the data. More feature of the neighborhoods can be included that can potentially affect the guests booking decisions.

4.0 Dependencies and Risks

Risk	Likelihood	Impact
Majority of the properties are available for booking for almost 200 days of the year in a tourist destination like New York City.	Low	Airbnb must have excess capacity in some neighborhoods which needs to be corrected to reduce the cost of revenue.
Guests consider only price while booking a room in Airbnb	Low	There may be other factors of the neighborhood which affects the decision of the guests and gaining knowledge of that will help Airbnb to pick the best neighborhood for revenue generation.

Lack of understanding of the factors that affect guests booking decisions may lead to more wrong product development decisions and misuse of valuable company resources.	High	Faulty understanding will result in wrong product development, which will incur more expenses and will end up in failures.

5.0 Deliverable Timelines

Item	Major Events / Milestones	Description	Date
1.	Kick-off / Formal Request	*Introduction and meeting with project advisor*	*07/04/2022*
2.	Analysis Plan & Data Finalization	*Determining the research objectives*	*07/15/2022*
3.	Data Exploration	Conduct exploratory data analysis	*07/22/2022*
4.	Model Building	Building various models as per analysis plan	*08/05/2022*
5.	Governance	Risk management to identify possible sources of biases in the research and finding ways to mitigate those biases.	*08/05/2022*
6.	Documentation	Starting documentation of the analysis report	*08/12/2022*
7.	Peer Feedback	Reviewing the works of peers and provide feedback	*08/12/2022*
8.	Presentation	Presenting the analysis report to the stakeholders	*08/22/2022 to 08/24/2022*
9.	Portfolio	Uploading the analysis in GitHub	*08/26/2022*

Reference:

Airbnb - statistics & facts. (n.d.). *Topic: Airbnb*. Statista. Retrieved July 14, 2022, from https://www.statista.com/topics/2273/airbnb/

Airbnb Inc. (2022). *Topic: Airbnb Annual Report 2021*. Retrieved July 14, 2022, from https://d18rn0p25nwr6d.cloudfront.net/CIK-0001559720/2a413af0-3429-4317-9d3c-a71f2d6d2683.pdf