

Capstone Project Submission

Short summary of the Capstone project , its components, problem statement, approaches and conclusions.

Airbnb stands for "Air Bed and Breakfast, it is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Problem Statement : Our aim is to analyze or predict the data in such a way that it will help new airbnb users to book for the first time . The accurate analysis helps to decrease the average time required to book by sharing more personalized recommendations and also in better forecasting demand by analyzing different features like country information , room type, price, neighborhood etc .

Approaches :

The problem is divided into multiple steps

1. Data Wrangling - to perform eda we imported different libraries like numpy , pandas , matplotlib and seaborn After importing the data we have done data wrangling to know more about data , by cleansing the data from the noise or flawed missing elements, datatype of columns, unique values, outlier treatment .
2. Descriptive statics- We used different types of graphs/plots to describe our data much better with the help of plots like histogram , pie chart, distribution plot, box plot , bar graphs and scatter plot to find the relationship between various variables.
3. EDA- In eda we visualize and analyze about
 - Different areas (neighborhood and boroughs) , how hosting is divided among these boroughs, the top neighborhoods and to which borough they belong , price distribution of each borough.
 - different room types and number of reviews.
 - Host and also some of the top and busiest hosts in our dataset.
 - Prices and how number of reviews , minimum nights are affecting price and also relation between price and availability 365.
 - Availability_365 - Manhattan and Brooklyn are heavily booked

because we found out that these two have availability less than 90 days.

Conclusions:

- 1) Among all 5 boroughs Manhattans and Brooklyn are the ones who have the most rentals. Manhattan is the most famous tourism borough and that is one of the reasons for most rentals.
- 2) Private rooms are the most reviewed across NYC.
- 3) Entire home provides great space as compared to others .
- 4) Most rentals are there for entire homes or private rooms and just very few in case of shared rooms. Shared room might not be much preferred by visitors.
- 5) Number of neighborhoods is highest in Queens and lowest in Manhattan still Manhattan has the most no. of rentals.
- 6) Due to the tourism factor the average price of rentals is higher in Manhattan and after Manhattan Brooklyn follows it.
- 7) Rentals whose availability is less than 90 days are generally the ones with lower prices.
- 8) Private room has the lowest minimum nights average and is also cheaper as compared to the entire room.

GitHub Repo link, Google Drive Link , Tableau link

Github Link:- https://github.com/smriti192002/AirbnbNYC_2019_Analysis

Google Drive Link -
<https://drive.google.com/drive/folders/17MtdPgMILVggQYKWh-ShWThq7l8OBns4?usp=sharing>

Tableau link -
https://public.tableau.com/app/profile/smriti7268/viz/AirBNBDataAnalysis_16568629230620/AirBnBAnalysis

Team Member's Name, Email and Contribution:

- | | |
|-------------------|---|
| 1. Smriti Bhattra | Email- smritib954@gmail.com |
| 2. Aman Gupta | Email- gupta4327@gmail.com |
| 3. Kiran Mahara | Email- kiranmahara20@gmail.com |

Contribution

1) Smriti Bhattra

1. Data Wrangling .
2. Treating null values .
3. Analyzing best features for analysis.
 - Room Types feature analysis
 - Price analysis
 1. deciding threshold for price by analyzing the price distribution graph.
4. minimum nights analysis and how it affects price
5. creating different graphs for data visualization in Tableau

1) Aman Gupta

Aman Gupta

1. Data Wrangling
 1. Analyzing relationship between null values of Last Review and Reviews per month column.
 2. Analyzing best features for analysis.
2. Area Analysis
 1. Borough analysis.
 2. Neighborhood analysis.
 3. Latitude – Longitude analysis
4. Code structuring

5. Dashboard Creation in Tableau

Kiran Mahara

1. Data Wrangling

1. Treating Null values.
2. Analyzing useful features for analysis.
2. Host analysis
3. Availability_365 analysis
 1. Categorizing numerical values of availability_365 into categories.
 2. Zero availability rental analysis.
4. Deciding best suited graphs to convey insights
5. Creating different graphs for data visualization in Tableau