**Data Analytics- Lodging Data**

[**Introduction**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#introduction)

* [**Scenario**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#scenario)
* [**Phase 1: Ask**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-1-ask)
  + [**Business objective**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#business-objective)
  + [**Business task**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#business-task)
  + [**Audience**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#stakeholders)
* [**Phase 2: Prepare**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-2-prepare)
  + [**Where is Data located**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#where-is-data-located)
  + [**How is the Data Organized?**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#how-is-the-data-organized)
  + [**Challenges with the data**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#challenges-with-the-data)
* [**Phase 3: Data Process**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-3-data-process)
  + [**What tools are you choosing and why?**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#what-tools-are-you-choosing-and-why)
  + [**Review of Data**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#review-of-data)
  + [**Collect Data(Data wrangling)**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#collect-datadata-wrangling)
* [**Phase 3: Data Cleaning**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-3-data-cleaning)
* [**Phase 4: Data analysis**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-4-data-analysis)
* [**Phase 5: Data findings with visualizations**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-5-data-findings-with-visualizations)
  + [**Visualization:1->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization1-)
  + [**Visualization:2->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization2-)
  + [**Visualization:3->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization3-)
  + [**Visualization:4->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization4-)
  + [**Visualization:5->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization5-)
  + [**Visualization:6->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#visualization6-)
* [**Phase 6: ACT**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#phase-6-act)
  + [**Key Takeaways->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#key-takeaways-)
  + [**Recommendations->**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#recommendations-)
  + [**Additional analysis for future**](https://rstudio-pubs-static.s3.amazonaws.com/878454_86d8b957d072434b8c865ab0f4c09908.html#additional-analysis-for-future)

**Introduction**

**This is a case study that I conducted to solve for my personal Project. It involves a lodging dataset which is given by Big-D Constructions, Utah. However, the company wants to know a story of the data provided. The management believes that doing this, they will be able to start a new marketing strategy that will be key for future growth of the company. As I have learned I will follow the steps of the data analysis process: ask, prepare, process, analyze, share and act. However, since act step is for executives to decide, I will not cover that step in this solution.**

**Scenario**

The given data is having the details of Airbnb. The data set includes the Property ID, Property name, Host ID, Host name, the Coordinates, City, Neighborhood, Price, Room type, Minimum nights of stay, Number of Reviews and Reviews per month.

The Team will expect the analysis for the following:

* To increase profitability
* To check the performance of each host in different cities
* To compare the price of the properties with different cities

**Phase 1: Ask**

**Business objective**

**To increase profitability by increasing the reviews of the property.**

**Business task**

**I as an analyst going to analyze the given data to answer the following questions:**

* **Average of Price, Reviews, Availability for different types of rooms in each city.**
* **Average of Price, Reviews, Availability for different types of rooms in each neighborhood.**
* **Performance of a host.**
* **Count of room types in each city.**
* **Average of Availability\_365 of the property in Neighborhood.**

**Audience**

**The Audience for this project:**

* The Recruiting team of Big-D Constructions, Utah.

**Phase 2: Prepare**

**Where is Data located**

The data used for this analysis were obtained from the Big-D Constructions for recruiting purpose. Here is the link for the data [Link](file:///C:\Desktop\Data%20Analyst%20Study%20Material\Data%20Analysis%20with%20Excel\Excel%20project\Bikesales%20project\Link).

**How is the Data Organized?**

The data is organized in csv file. The files consist of 17 columns containing information related to Property ID, Property name, Host ID, Host name, the Coordinates, City, Neighborhood, Price, Room type etc.

**Credibility of the Data**

The data is comprehensive as it consists of data related to Airbnb and it is just a sample of the data.

**Challenges with the data**

* There are some problems with the data as well. Most of the problems (duplicate records, missing fields, etc.) can be dealt with by data cleaning.

**Phase 3: Data Process**

**What tools are you choosing and why?**

For this project I choose MS-Excel in order to prepare, process, clean, analyze and create the visualizations.

**Review of Data**

In order to get an overview, the data was reviewed in terms of understanding of the consent of variables, data formats and data integrity.

**Data review involved the following:**

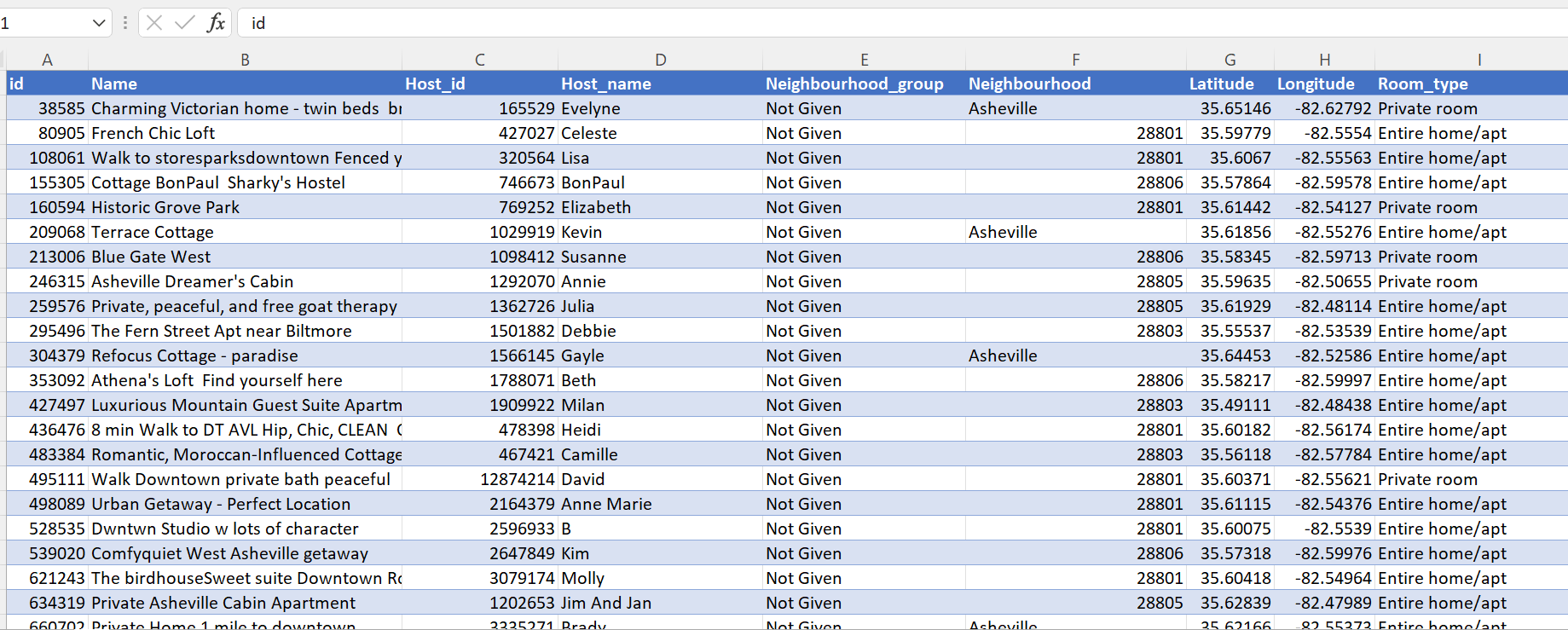
* Checking column names.
* Checking for missing values.
* Checking of white spaces.
* Checking of duplicate records.
* Other data anomalies.

**However the review of the data revealed several problems:**

* Property names with special characters and white spaces.
* Dates given was not in correct format.
* Missing data

**Collect Data(Data wrangling)**

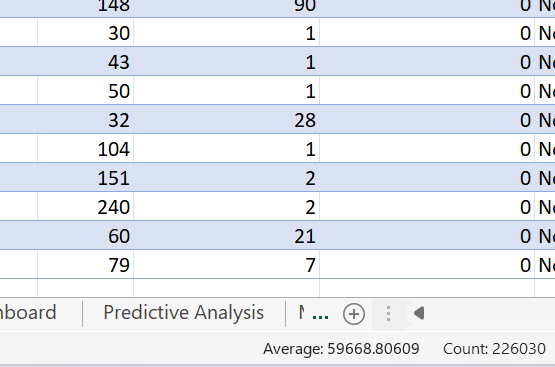
* At this point, I had to convert CSV files in .xlsx format and uploaded into MS-Excel.



**Data validation**

* Check column names to ensure we can join all the data. Compare column names for each of the files.
* Calculate the total number of records in the file. It is 226030 rows and 17 columns in

Total.



**Phase 3: Data Cleaning**

**We need to start data cleaning now.**

* Converted the dataset in to a table which makes cleaning easy using Power Query Editor.
* Removed the special characters in the Name column using Power Query Editor and save into a new sheet.
* Replaced all the blanks as “Not Given” using Power Query Editor.
* Converted the date into Short Date format.

**Phase 4: Data analysis**

**Data validation->**

**Upload the trip\_clean.csv file on-board and check data validation if everything is ok.**

* **Analysis-1: Average of Price, Reviews, Availability for different types of rooms in each city**
* **Analysis-2: Average of Price, Reviews, Availability for different types of rooms in each neighborhood.**
* **Analysis-3: Performance of a host.**
* **Analysis-4: Count of room types in each city.**
* **Analysis-5: Average of Availability\_365 of the property in Neighborhood**
* **Analysis-6:**

**Phase 5: Data findings with visualizations**

**Visualization:1->**

**This is the first visualization which shows the average of price, average of number of reviews and availability of the property in different city sliced using different types of room. This helps in getting a clear picture of a property’s popularity in the particular city.**

**Visualization:2->**

**This visualization shows the average of price, average of number of reviews and availability of the property in different neighborhood sliced using different types of room. This helps in getting a clear picture of a property’s popularity in the particular neighborhood.**

**Visualization:3->**

**Now, let’s see what is the performance of a host based on how many properties he/she dealing with , number of reviews, average availability of the property. This gives the visualization of the host’s performance in the business.**

**Visualization:4->**

**On my fourth plot, I tried to observe the count of rooms based on the types like Entire apartment, Shared room, Hotel room and Private room. This visualization has been sliced based on the city. So, here we can view the total count of rooms in each city based on the room types.**

**Visualization:5->**

**Visualization:6->**

**Phase 6: ACT**

**Key Takeaways->**

**Recommendations->**

**Additional analysis for future**