**Data Visualization (Tableau and Power BI) by TalentLabs**

**Final Project (Total Marks 100)**

Dear Students,

This is your final project in which you have to choose your data set, understand and explore data insights using TABLEAU or POWER BI and output the final report. Your final dashboard should be in proper and presentable format. Below are the guidelines to follow. Go through the questions and prepare your answers accordingly.

**Data Set Selection:**

Below are some datasets to be used in this project. You can choose any of them. Before finalizing, make sure that you understand it. You should be able to extract maximum of information out of it using visuals and data insights.

  

**Part 1**

**Data Cleaning (7.5 Marks)**

Check your chosen dataset if it needs any data cleaning. Go through all the fields, check for any null values or incorrect data types. Answer the following questions.

1. Share your findings for data cleaning.
2. How are you dealing with null values?
3. Do you think some item data types need to be changed? Why do you think so.

We chose Airbnb dataset for this data visualization project. During data cleaning, we will deal with some problems such as handling missing values in our dataset, changing our data column to appropriate data type, and we may need to remove irrelevant column(s) for data analysis.

1. **Handling Missing Values**: Certain columns have missing entries as below.

|  |  |  |
| --- | --- | --- |
| **Column** | **Row Counts of Missing Entries** | **How to solve?** |
| Review Scores Rating (bin) | 8,309 | Create separate columns to categorize the rows where their values are missing. Then, replace those missing entries as zero. |
| Review Scores Rating | 8,309 |
| Beds | 85 |
| Host Since | 3 | Remove those 3 rows |
| Property | 3 | Replace it with the value “Other” |

Another problem we should deal with during data cleaning is **Unsuitable Data Types**. If left unhandled, there will be some issues when we deal with aggregation like SUM, AVERAGE or COUNT in Power BI.

Below are some important columns we need to change them to appropriate data type:

|  |  |  |
| --- | --- | --- |
| **Column** | **Correct Data Type** | **Reason to change** |
| Host Since | datetime | to analyze trends over time properly. |
| Review Scores Rating (bin) | Integer | To deal with aggregation like SUM, AVERAGE, MIN, MAX, etc |
| Review Scores Rating | Integer |
| Beds | Integer |
| Price | Integer |
| Zipcode | string - categorical | To be treated as categorical variable |
| Neighborhood | string – categorical |
| Property Type | string – categorical |
| Room Type | string – categorical |

**Redundant or Irrelevant Column:** The 'Number of Records' column contains only the value 1 for all entries, making it redundant for any analysis. And that’s why we will remove this column.

**Part 2**

**Understanding your dataset (15 Marks)**

1. Why have you chosen this dataset?
2. Explain your dataset in detail.
3. What key performance indicators you can identify of guests?
4. What new information and indicators can be drawn through this dataset?
5. What analysis can make this a purpose full report?
6. What kind of modeling you are planning to use? Please elaborate reasons to choose model.

The reason I chose this Airbnb dataset is that it reflects a rich variety of information about Airbnb listings, which can offer valuable insights into the sharing economy, real estate trends and rental market prices.

The raw dataset of Airbnb contains 30,478 rows and 13 columns. The dataset contains information related to Airbnb listings, including:

|  |  |  |
| --- | --- | --- |
| **Data Column** | **Data type** | **Description** |
| Host Id | Integer | ID on Airbnb |
| Host Since | Date | Registration date on Airbnb |
| Name | String – Categorical | The property name |
| Neighborhood | String – Categorical | Neighborhood information |
| Property Type | String – Categorical | Type of property |
| Room Type | String – Categorical | Type of room |
| Zipcode | Integer | The zipcode of the property |
| Beds | Integer | Number of beds in the property |
| Number of Reviews | Integer | Total number of reviews |
| Price | Integer | Price of the property per night |
| Review Scores Rating | Integer | Rating of the property based on reviews |
| Number of Records | Integer | Total number of records in the dataset |

Key performance indicators for guests may involve examining the average price per room type and neighborhood, assessing the average number of reviews per listing as a proxy for occupancy rate, and considering the Review Scores Rating as an indicator of guest satisfaction.

In terms of new insights from the dataset, one can derive potential occupancy rates based on the number of reviews, explore seasonality in booking rates using 'Host Since' dates, and potentially uncover links to pricing trends. Additionally, analyzing the influence of property type, room type, and location on customer satisfaction and price can provide valuable information.

To make the report purposeful, conducting correlation analysis between price, number of reviews, and review scores, performing comparative analyses of different neighborhoods and room types regarding price and reviews, and employing time-series analysis to understand trends and seasonality effects would be beneficial.

Regarding modeling, regression models could be applied to predict factors influencing price or satisfaction scores, time-series analysis can help identify trends in prices and reviews over time, and cluster analysis may be useful to group similar listings and identify unique characteristics. The choice of these models is driven by their suitability for capturing and interpreting different aspects of the dataset.

**Part 3**

**Identify focused Area of Analysis (15 Marks)**

In this part, you need to identify which fields/columns of your selected dataset, you will focus more to analyze. Please answer below questions.

1. Identify and select which fields/columns your will focus your analysis?
2. Please provide reason for choosing these fields/columns? Why do you think they should be more focused?
3. What is your current understanding of these selected fields/column?
4. What conclusion you can draw after drilling down these columns/fields?
5. Which other fields can be directly or indirectly impacted with your conclusion?

|  |
| --- |
| **Notes for Part 4 and Part 5**  For the Part 4 and Part 5 questions, please follow below instructions before you proceed.     * Please understand your chosen dataset carefully. * Do some homework to identify what information is currently existing and what information you have to create. * Note down key findings from your dataset as a draft. * Make a plan of how to show or display your findings. * Identify what other visuals can be created using formulas. * Choose your platform (Tableau or Power BI) wisely. * Make sure you have cleaned your data before going into visualization. * Make sure your visuals are presenting different content. * Make sure you are using different visual types. One visual type should not be repeated more than 2 times. (Card visuals should not be repeated more than 3 times). * Make sure your workbook/visuals/dashboard is in proper format, title and labeling. * Your final submission should include your Tableau /Power BI workbook and your datasets. |

**Part 4**

**Preparing for Dashboard Report (20 marks)**

1. What platform you are using for dashboard report? Explain in 5 points that why you have given preference to your chosen platform over others?
2. Can you draft a sketch of how do you want your dashboard to look like?
3. Where do you want your audience to put attention to? How are you planning to do that?
4. How are you planning to create data story through your visuals?
5. Can you elaborate the narrative you are using for data story?
6. How are you ensuring that your visuals are interactive?
7. What formulas/DAX/Calculated fields you have used? What visuals you have used for them?
8. How did you verify your formulas if they are returning correct values?

**Part 5**

**Dashboard Report Execution (20 Marks)** Data Cleaned **5 Marks**

Visuals Formatting **5 Marks**

Dashboard Layout **5 Marks**

Data Insights **5 Marks**

Prepare a dashboard report using TABLEAU or POWER BI platform. Please attach your workbooks along with your datasets.

**Part 6**

**Final Report (12.5 Marks)**

1. Briefly explain your dashboard workbook layout and legends.
2. Please explain what information your visuals are sharing?
3. Please explain what problems and challenges you have faced during this project.
4. What are your key learning points during this assignment? Please mention any three.
5. How are you going to apply these skills in your daily routine tasks? Briefly mention how this project will help you?

**Part 7**

**Create Problem Statement (10 Marks)**

Now you have analyzed your dataset, created visuals and developed a dashboard. You should be able to derive all important and critical information from your analysis by now.

Please answer below questions based on your final analysis.

* 1. Please clearly state and explain what problem or problems you have drawn from your analysis?
  2. What are the reasons of these problems? How did you identify problem root cause?
  3. What actions can be recommended through your analysis as a remedy of problems? Please mention remedy against each problem.
  4. What factors are putting negative and positive impact on key performance indicators? Please list down.

**Submission**

You are required to submit a report in docx or ppt format that cover everything above. You don’t need to answer the questions one by one like in exam papers. Instead, your submission should be holistic document that cover all the 7 parts above.

Remember to share the access to the dashboard before you submit it, so our assignment graders would be able to see and give feedback on your great creations.