Assessment for Data Analyst Role (Tableau, Power BI)

**Part 1: Theoretical Knowledge**

1. Understanding Data Visualization

Question: **Explain the importance of data visualization in data analysis. What are the key principles of effective data visualization?**

* Data visualization is crucial in data analysis because it simplifies complex information, making it easier to understand and communicate. It helps identify patterns, trends, and outliers, which can lead to better decisions.

Key principles of effective data visualization include:

1. **Clarity**: Make it easy to understand.
2. **Accuracy**: Represent the data truthfully.
3. **Relevance**: Focus on important data.
4. **Consistency**: Keep design elements uniform.
5. **Audience Awareness**: Tailor it to the audience's needs.
6. **Aesthetics**: Make it visually appealing without distracting from the data.
7. **Storytelling**: Guide the viewer through the data logically.

2. Tableau Basics

Question: **What are the main components of Tableau? Describe the process of creating a basic dashboard in Tableau.**

**Tableau main components**

1. **Data Source**: Where your data comes from (e.g., Excel, SQL).
2. **Sheets**: Workspaces where you create individual visualizations.
3. **Dashboard**: A collection of multiple visualizations in one view.
4. **Story**: A sequence of dashboards or sheets that tell a narrative.
5. **Marks**: Controls how data is displayed (color, size, etc.).
6. **Filters**: Used to show specific parts of your data.
7. **Pages**: Breaks down a view into a series of visualizations.

**Creating a Basic Dashboard in Tableau step-by-step process.**

1. **Connect to Data**: Import your data into Tableau.
2. **Create Sheets**: Drag and drop data fields to make charts or graphs.
3. **Build a Dashboard**: Combine your sheets into a single view.
4. **Customize**: Arrange, resize, and add interactive elements.
5. **Share**: Save or publish the dashboard to share with others.

3. Power BI Fundamentals

Question: **Discuss the main features of Power BI. How does Power BI differ from Tableau in terms of functionality and use cases?**

**Main Features of Power BI**

* Power BI connects to various data sources like Excel, SQL & Azure and allows easy data transformation Cleans and models using **Power Query**. It offers interactive visualizations for dynamic data exploration and lets you create dashboards and reports for sharing insights. **DAX** enables advanced calculations, while AI features and collaboration tools enhance analysis and teamwork across the organization.

Here's a comparison of **Power BI and Tableau**:

|  |  |  |
| --- | --- | --- |
| Feature | Power BI | Tableau |
| Ease of Use | User-friendly, especially for those familiar with Microsoft products. | More complex, offering advanced customization options. |
| **Data Handling** | Best within Microsoft ecosystem (e.g., Excel, Azure). | Superior for handling large datasets and complex data sources. |
| **Visualization** | Good for standard visualizations and reports. | Excellent for detailed, high-quality visualizations. |
| **Cost** | More affordable, with a free version (Power BI Desktop). | Higher cost, often better suited for large enterprises. |
| **Integration** | Seamless integration with Microsoft services like Teams and SharePoint. | Integrates well with a wide range of platforms and data sources. |
| **Use Cases** | Ideal for business users needing basic reporting and dashboarding. | Preferred by data professionals for deep data analysis and exploration. |

**Part 2: Practical Application**

4. Data Cleaning and Preparation

**Problem Statement**: Given a dataset with missing values and inconsistencies, clean and prepare the data for analysis.

 Here I attached My worked Project. Kindly check.

### Documentation

1. **Import Libraries**: Import necessary libraries such as Pandas and NumPy for data manipulation and handling.
2. **Load the Dataset**: Load the dataset into a Pandas Data Frame to start the cleaning process.
3. **Explore the Data**: Inspect the dataset to understand its structure, data types, and missing values.
4. **Handle Missing Values**: Identify and address missing values by either dropping them or filling them with appropriate values.
5. **Fix Data Inconsistencies**: Correct inconsistencies by removing duplicates, adjusting data types, standardizing formats, and handling outliers.
6. **Verify and Save**: Ensure that the data is cleaned properly and save the cleaned dataset for further analysis.

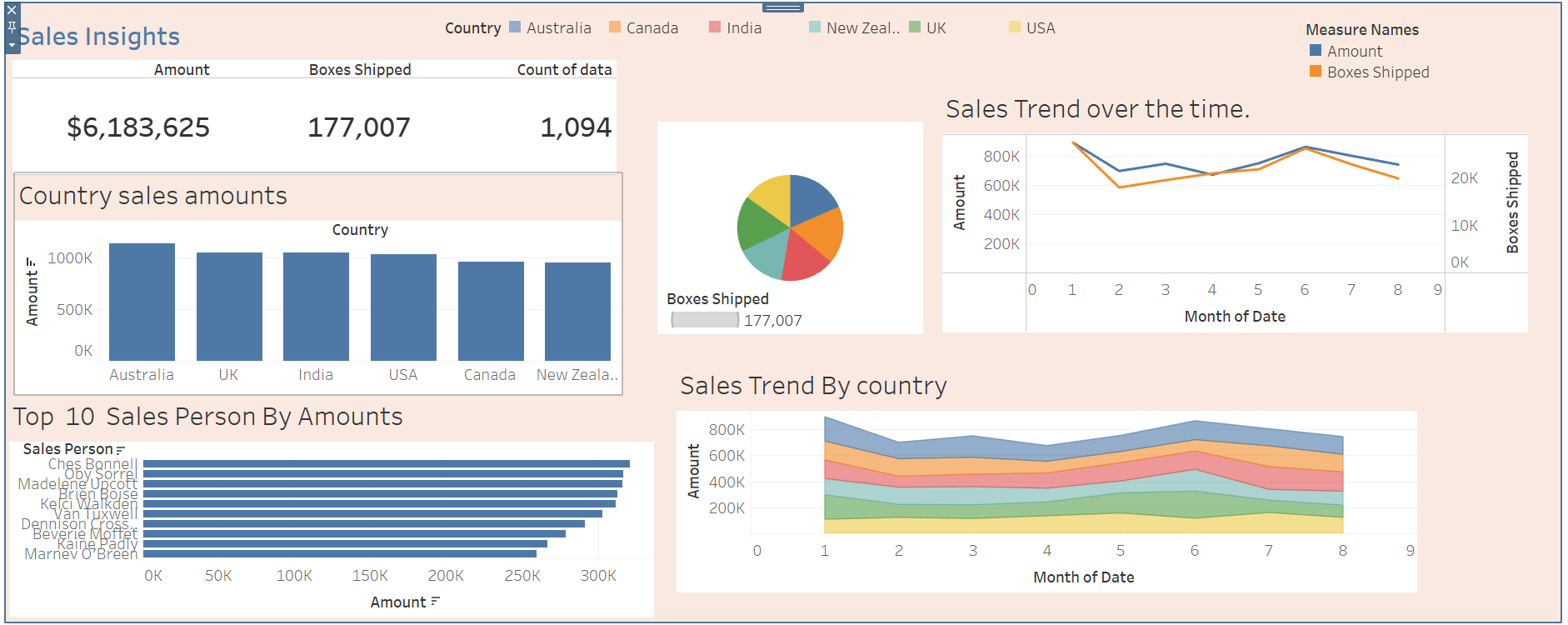
By following these steps, you will ensure that your dataset is well-prepared for analysis, with missing values and inconsistencies appropriately addressed.

5. Tableau Visualization

**Problem Statement:** Create an interactive sales dashboard in Tableau using the provided sales dataset. The dashboard should include key metrics such as total sales, sales by region, and sales trends over time.

Here I share the sales report dashboard using Tableau.

Tableau public Dashboard link: <https://public.tableau.com/views/SalesReport_17253553369450/Dashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>



**Steps to Create a Sales Dashboard in Tableau**

1. **Load the Data**:
   * Open Tableau and connect to your sales dataset.
2. **Create Visualizations**:
   * **Total Sales**: Create a text Tables showing total sales amount, box shipped and counts
   * **Sales by Region**: Make a bar chart displaying sales by country or region.
   * **Sales Trends**: Create a line chart showing sales over time.
3. **Build the Dashboard**:
   * Combine your visualizations onto a single dashboard canvas.
   * Arrange and resize the charts to fit well.
4. **Add Filters**:
   * Include filters for regions and dates to allow users to interact with the data.
5. **Customize and Finalize**:
   * Add a title, labels, and adjust colors for clarity.
   * Test the interactivity.
6. **Publish and Share**:
   * Save and publish your dashboard to share with others.

6. Power BI Report

**Problem Statement:** Develop a report in Power BI to analyze customer feedback data. The report should highlight customer satisfaction levels, common issues, and trends over time.

Here I attached Redmi 6 Customer Feedback Report:



Power Bi Dashboard Link: <https://app.powerbi.com/view?r=eyJrIjoiNDdhZjQ3MmEtZGU5MC00MjhiLTg0YjktZmYzNjk4OWY2NjJjIiwidCI6ImM2ZTU0OWIzLTVmNDUtNDAzMi1hYWU5LWQ0MjQ0ZGM1YjJjNCJ9>

**Steps to Develop a Customer Feedback Report in Power BI**

1. **Load the Data**:
   * Import your customer feedback dataset into Power BI.
2. **Create Visualizations**:
   * **Customer Satisfaction**: Use a pie chart to show satisfaction levels.
   * **Common Issues**: Create a bar chart highlighting the most common issues.
   * **Trends Over Time**: Use a line chart or timeline to display feedback trends over time.
3. **Add Slicers**:
   * Include slicers to filter the data by criteria such as date, region, or product type.
4. **Design the Report**:
   * Arrange the visualizations neatly on the report page.
   * Add titles and labels to make the report clear and easy to understand.
5. **Test Interactivity**:
   * Ensure slicers and visualizations interact correctly and reflect changes in the data.
6. **Publish and Share**:
   * Save and publish your Power BI report, then share it with stakeholders.

7. Statistical Analysis

**Problem Statement:** Perform a statistical analysis on a given dataset to identify significant trends and correlations. Provide a summary of your findings.



Here I attached some statistical analysis for my projects.

**Steps for Statistical Analysis**

1. **Load the Data**:
   * Import your dataset into a tool like Python (with Pandas) or Excel.
2. **Perform Correlation Analysis**:
   * Use correlation methods (e.g., Pearson) to identify relationships between variables.
3. **Conduct Hypothesis Testing**:
   * Apply hypothesis tests (e.g., t-test) to determine if observed trends are statistically significant.
4. **Visualize the Results**:
   * Create visualizations like scatter plots or heatmaps to display correlations and trends.
5. **Summarize Findings**:
   * Write a clear summary of the key trends and correlations you discovered.

8. Predictive Analytics

**Problem Statement:** Build a predictive model to forecast sales for the next quarter using historical sales data. Explain the steps taken and the rationale behind your model choice.

 for this project is predict the prize for the particular category (eg. Seqf, model, mileage). In the result Random Forest Regression accuracy is 95%

**Steps to Build a Predictive Sales Model**

1. **Load the Data**:
   * Import your historical sales data using Python (e.g., Pandas).
2. **Data Preprocessing**:
   * **Clean the Data**: Handle missing values, remove outliers, and format dates.
   * **Feature Engineering**: Create new features like monthly sales or seasonal indicators.
   * **Split the Data**: Divide the data into training and testing sets (e.g., 80/20 split).
3. **Choose a Model**:
   * Select a model (e.g., Linear Regression, Lasso, Decision Tree Regression) based on the nature of your data.
4. **Build the Model**:
   * Train the model using the training data.
   * Tune hyperparameters to optimize performance.
5. **Evaluate the Model**:
   * Test the model on the testing set and evaluate metrics like RMSE or MAE.
   * Compare predicted values with actual sales to assess accuracy.
6. **Cardheko price Prediction**:
   * Use the model to predict cardheko price for the category.
   * Plot the forecasted values alongside historical data for comparison.

**Part 4: Scenario-Based Questions**

9. Real-World Problem Solving

Question: Imagine you are given a large dataset with customer transactions. How would you approach the task of identifying key customer segments and their behaviors? Describe the steps and tools you would use.

**Approach to Identifying Key Customer Segments and Behaviors**

1. **Understand the Objectives**:
   * **Action**: Clarify the business goals for segmentation, such as improving marketing strategies or identifying high-value customers.
2. **Load and Explore the Data**:
   * **Tools**: Use Python (Pandas) or R.
   * **Action**: Load the dataset and perform initial exploration to understand the range of features available, such as purchase history, customer demographics, and behavior patterns.
3. **Preprocess the Data**:
   * **Tools**: Python (Pandas, Scikit-learn) or R.
   * **Action**: Clean the data by addressing missing values, standardizing formats, and normalizing features like spending amounts and frequency of transactions.
4. **Feature Engineering**:
   * **Tools**: Python (Pandas) or R.
   * **Action**: Create new metrics like RFM (Recency, Frequency, Monetary) scores or customer lifetime value (CLV) to enhance the segmentation process.
5. **Apply Segmentation Techniques**:
   * **Tools**: Python (Scikit-learn) or R.
   * **Action**: Use clustering algorithms (e.g., K-Means, DBSCAN) to identify natural groupings of customers based on the features you've prepared. You could also consider hierarchical clustering or dimensionality reduction techniques like PCA to simplify the data before clustering.
6. **Analyze and Profile Segments**:
   * **Tools**: Python (Matplotlib, Seaborn) or R.
   * **Action**: Analyze each segment to identify key characteristics, such as demographics, purchasing behavior, and engagement levels. Use visualizations to highlight the differences between segments.
7. **Validate and Refine the Segments**:
   * **Tools**: Python (Scikit-learn) or R.
   * **Action**: Validate the segments by checking for consistency over different time periods or by applying the model to a new dataset. Refine the segmentation model if necessary, using feedback from stakeholders or additional data.
8. **Communicate the Insights**:
   * **Tools**: Tableau, Power BI, or Python (Matplotlib/Seaborn).
   * **Action**: Create visual reports and dashboards to present the segments and their behaviors to stakeholders. Highlight actionable insights, such as which segments to target for specific marketing campaigns or loyalty programs.

This approach will enable you to systematically uncover key customer segments and understand their behaviors, providing actionable insights for better decision-making in areas like marketing, sales, and customer retention.

10. Data-Driven Decision Making

Question: A company wants to launch a new product and has collected survey data on customer preferences. How would you use this data to help the company make an informed decision? Outline your approach.

**Approach to Using Survey Data for a Product Launch Decision**

1. **Understand the Objective**:
   * Clarify the company’s goals for the product launch (e.g., target market, key features).
2. **Load and Clean Data**:
   * Use tools like Python or R to clean the survey data, handling missing values and inconsistencies.
3. **Analyze Customer Preferences**:
   * Identify key trends and preferences, such as which features customers value most and their price sensitivity.
4. **Segment Customers**:
   * Group customers into segments based on their preferences and demographics using clustering techniques.
5. **Predict Outcomes**:
   * Build models to predict how different features or configurations might influence customer satisfaction or purchase decisions.
6. **Conduct Scenario Analysis**:
   * Test various product scenarios (e.g., different features or prices) to understand their potential impact on sales.
7. **Make Recommendations**:
   * Present actionable insights, such as which features to prioritize and which customer segments to target.
8. **Plan for Monitoring**:
   * Suggest a post-launch plan to monitor sales and customer feedback, enabling continuous improvement of the product.

This approach ensures that the company makes data-driven decisions that align with customer preferences, increasing the likelihood of a successful product launch.

-------------------------------------------------------------------------------------------------------------------