

SNS GROUPS

Assessment for Data Analyst Role (Tableau, Power BI)

Objective:

Evaluate the candidate's proficiency in data analysis, visualisation, and reporting using Tableau and Power BI.

Part 1: Theoretical Knowledge

1. Explain the importance of data visualisation in data analysis. What are the key principles of effective data visualisation?

Data visualisation is a critical component of data analysis because it transforms complex data sets into accessible, interpretable visual formats. Here's why data visualisation is important and the key principles of effective visualisation:

Importance of Data Visualization

1. **Enhanced Understanding:** Visualisations help in grasping complex data patterns, trends, and correlations that might be difficult to detect in raw data. They turn abstract numbers into a more intuitive form, making the data more comprehensible.
2. **Quick Insights:** Effective visualisations can quickly convey insights and key messages. For example, a well-designed chart can immediately show a trend or highlight outliers, saving time and reducing the need for detailed analysis.
3. **Improved Communication:** Visualisations make it easier to communicate findings to various stakeholders, including those without a technical background. Clear visuals help in making data-driven decisions and persuading others of your findings.
4. **Pattern Recognition:** Humans are better at recognizing patterns and anomalies in visual representations than in raw numerical data. Visualisations can highlight these patterns, making it easier to make informed decisions.
5. **Storytelling:** Good visualisations can tell a story by presenting data in a logical flow. This helps in illustrating how data points are interconnected and what they signify in a broader context.

Key Principles of Effective Data Visualization

1. **Clarity and Simplicity**
 - **Minimise Clutter:** Remove unnecessary elements that do not contribute to the understanding of the data. Use white space effectively to avoid overwhelming the viewer.
 - **Clear Labels:** Ensure that all axes, legends, and labels are clearly marked and easy to read. The visualisation should be self-explanatory.
2. **Accuracy**
 - **Correct Representation:** Ensure that the data is represented accurately without distortion. Avoid misleading scales or manipulations that could misinterpret the data.

- **Proper Scaling:** Use appropriate scales and avoid manipulative practices like truncating axes to exaggerate differences.
- 3. **Appropriate Chart Selection**
 - **Match Chart Type to Data:** Choose the type of chart that best represents the data. For example, use line charts for trends over time, bar charts for comparisons, and scatter plots for relationships between variables.
 - **Avoid Overuse:** Avoid overloading a single visualisation with too much information. Break data into multiple visuals if necessary.
- 4. **Consistency**
 - **Uniform Design:** Use consistent colours, fonts, and design elements throughout your visualisations. This helps in maintaining coherence and avoiding confusion.
 - **Standardised Elements:** Apply standardised visual elements like scales and symbols to facilitate easy comparison and interpretation.
- 5. **Focus on the Audience**
 - **Understand the Audience:** Tailor the complexity and detail of the visualisation to the audience's level of expertise and interest.
 - **Interactive Elements:** For digital platforms, consider adding interactive elements like filters or drill-downs to allow users to explore the data further.
- 6. **Effective Use of Colour**
 - **Meaningful Colour Choices:** Use colours to highlight key points or differences. Avoid using too many colours, which can be distracting.
 - **Colour Blindness Consideration:** Ensure that colour choices are accessible to people with colour vision deficiencies. Use patterns or textures as supplementary cues if necessary.
- 7. **Logical Flow and Storytelling**
 - **Narrative Structure:** Organise the visualisation to follow a logical sequence or narrative. Guide the viewer through the data with a clear beginning, middle, and end.
 - **Highlight Key Insights:** Emphasise the most important data points or trends to draw attention to the main message of the visualisation.
- 8. **Use of Annotations**
 - **Contextual Information:** Add annotations to provide additional context or explanations where needed. This can help in clarifying complex data points or trends.

2. Explain the importance of data visualisation in data analysis. What are the key principles of effective data visualisation?

Tableau is a powerful data visualisation tool that helps users create interactive and shareable dashboards. Its main components and the process of creating a basic dashboard are outlined below:

Main Components of Tableau

1. **Workbooks**
 - **Definition:** A Tableau workbook is a file that contains one or more dashboards, sheets, and data connections.
 - **Components:** It includes worksheets, dashboards, and stories.
2. **Worksheets**
 - **Definition:** A worksheet is where you create individual charts, graphs, and other visualisations.
 - **Components:** Includes data fields, visualisation types, and formatting options.

3. **Dashboards**
 - **Definition:** A dashboard is a collection of one or more worksheets and other elements (like images, text, and web content) arranged on a single screen.
 - **Components:** Allows for interactive features like filters, actions, and parameters.
4. **Stories**
 - **Definition:** A story is a sequence of sheets and dashboards that work together to convey a data-driven narrative.
 - **Components:** Composed of multiple "story points," each showing a specific view or set of views.
5. **Data Connection**
 - **Definition:** This is the link between Tableau and your data source. Tableau supports various data sources, including databases, spreadsheets, and web data connectors.
 - **Components:** Includes data connection settings, data import, and metadata management.
6. **Data Pane**
 - **Definition:** The area where you see all the data fields available for use in your visualisations.
 - **Components:** Includes dimensions (categorical data) and measures (quantitative data).
7. **Shelves and Cards**
 - **Definition:** Shelves and cards are areas where you drag and drop fields to build your visualisations.
 - **Components:**
 - **Columns Shelf:** Controls what goes on the X-axis.
 - **Rows Shelf:** Controls what goes on the Y-axis.
 - **Filters Shelf:** Controls which data is included in the visualisation.
 - **Pages Shelf:** Allows you to create animations or pagination effects.
8. **Marks Card**
 - **Definition:** A set of controls that help you define how data points are displayed.
 - **Components:** Includes options for colour, size, label, detail, and shape.
9. **Filters**
 - **Definition:** Filters allow you to control which data is displayed in your visualisation.
 - **Components:** Can be applied at various levels (e.g., data source, worksheet, dashboard).
10. **Parameters**
 - **Definition:** Parameters are dynamic values that can be used to modify calculations or control filters.
 - **Components:** Allow users to input values or select options that affect the data shown.
11. **Legends and Headers**
 - **Definition:** Legends provide information about the data in your visualisation, such as colour or size legends.
 - **Headers:** Display field names and values in tables or charts.

Process of Creating a Basic Dashboard in Tableau

1. **Connect to Data**
 - **Open Tableau:** Start Tableau and connect to your data source. You can connect to various types of data sources including Excel files, databases, or online services.
 - **Import Data:** Select the data you want to work with and set up the data connection.
2. **Prepare Data**
 - **Data Preparation:** Clean and transform your data if needed. You can create calculated fields, join multiple tables, and set up filters to refine the dataset.

3. **Create Worksheets**
 - **Drag and Drop:** Create individual visualisations by dragging and dropping fields from the Data Pane onto the Rows and Columns shelves.
 - **Choose Visualisation Type:** Select the appropriate chart type (e.g., bar chart, line chart, scatter plot) from the “Show Me” panel based on your data and analysis needs.
 - **Customise Visualisation:** Use the Marks Card to customise colours, labels, sizes, and other visual elements.
4. **Build the Dashboard**
 - **Create a New Dashboard:** Go to the Dashboard tab and create a new dashboard.
 - **Drag and Drop Worksheets:** Drag your created worksheets onto the dashboard canvas. Arrange and resize them as needed.
 - **Add Interactive Elements:** Add filters, parameter controls, and actions to make the dashboard interactive. This allows users to filter data or interact with different views.
5. **Format and Design**
 - **Layout and Design:** Adjust the layout of your dashboard to ensure it is visually appealing and easy to navigate. Add titles, text boxes, and images if needed.
 - **Optimise for Performance:** Ensure that the dashboard performs well by optimising data sources and minimising complex calculations.
6. **Publish and Share**
 - **Save the Workbook:** Save your Tableau workbook locally or on Tableau Server/Tableau Online if you have access.
 - **Publish Dashboard:** Publish your dashboard to Tableau Server or Tableau Online to share it with others. Set permissions as necessary.
7. **Iterate and Improve**
 - **Gather Feedback:** Collect feedback from users to identify areas for improvement.
 - **Refine Dashboard:** Make any necessary changes based on feedback and update the dashboard accordingly.

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

Power BI and Tableau are two leading tools in the business intelligence and data visualisation space, each with its own set of features and strengths. Here’s an overview of the main features of Power BI and how it differs from Tableau in terms of functionality and use cases:

Main Features of Power BI

1. **Data Connectivity**
 - **Wide Range of Connectors:** Power BI offers extensive connectivity options, including databases, cloud services, Excel files, and web data sources.
 - **Direct Query and Import Modes:** You can either import data into Power BI or use Direct Query mode to work with live data.
2. **Data Modeling**
 - **Power Query:** Provides a powerful and user-friendly interface for data transformation and cleansing.

- **Data Relationships:** Allows users to define relationships between tables and build complex data models.
- 3. **Visualisations**
 - **Variety of Charts and Graphs:** Power BI provides a rich set of built-in visualisations including bar charts, pie charts, maps, and more.
 - **Custom Visuals:** Users can download and integrate custom visuals from the Microsoft AppSource marketplace or create their own.
- 4. **Interactive Dashboards**
 - **Interactive Elements:** Dashboards are interactive and support slicers, filters, and drill-through capabilities.
 - **Real-time Data:** Power BI dashboards can be updated in real-time with data streaming and push datasets.
- 5. **Natural Language Queries**
 - **Q&A Feature:** Users can ask natural language questions about their data, and Power BI generates visualisations based on the query.
- 6. **Integration with Microsoft Ecosystem**
 - **Excel Integration:** Seamless integration with Excel for importing data and leveraging Excel formulas.
 - **Microsoft 365 Integration:** Works well with other Microsoft products like Azure, SharePoint, and Teams.
- 7. **Collaboration and Sharing**
 - **Power BI Service:** Allows for publishing, sharing, and collaborating on reports and dashboards online.
 - **Workspaces:** Organise and manage content in different workspaces within the Power BI service.
- 8. **Mobile Access**
 - **Mobile App:** Power BI provides mobile applications for viewing and interacting with reports and dashboards on-the-go.
- 9. **Security and Compliance**
 - **Row-Level Security:** Apply security rules to restrict data access based on user roles.
 - **Data Governance:** Comprehensive data governance and compliance features integrated with Microsoft's security framework.

How Power BI Differs from Tableau

1. **Cost**
 - **Power BI:** Generally more cost-effective, with a free version available and a lower-priced Pro version. Premium options are also available for larger organisations.
 - **Tableau:** Typically more expensive, with a higher cost for licensing and additional features. Offers a free Tableau Public version with limitations.
2. **Ease of Use**
 - **Power BI:** Often considered more user-friendly for beginners, particularly those familiar with Microsoft Office products. Its integration with Excel and other Microsoft tools is straightforward.
 - **Tableau:** While also user-friendly, Tableau is praised for its advanced visualisation capabilities and flexibility. It may have a steeper learning curve but offers more powerful customization options.
3. **Data Modeling and Transformation**
 - **Power BI:** Uses Power Query for data transformation and Power Pivot for data modelling. It integrates well with Microsoft products for complex data transformations.
 - **Tableau:** Provides Tableau Prep for data cleaning and transformation. It is highly regarded for its intuitive drag-and-drop interface and powerful data blending capabilities.

4. **Visualisation Flexibility**
 - **Power BI:** Offers a wide range of standard visualisations and custom visuals, but might be somewhat less flexible in highly complex visualisations compared to Tableau.
 - **Tableau:** Known for its flexibility and depth in creating sophisticated visualisations. It allows for detailed customization and advanced analytics features.
5. **Performance and Scalability**
 - **Power BI:** Performs well for standard business intelligence tasks and scales effectively within the Microsoft ecosystem. The performance of large data sets can be impacted by the complexity of the data model.
 - **Tableau:** Highly scalable and optimised for handling large and complex data sets. It is often preferred for enterprise-scale deployments requiring high performance.
6. **Deployment and Integration**
 - **Power BI:** Integrates seamlessly with other Microsoft products and services, making it an ideal choice for organisations already using Microsoft tools.
 - **Tableau:** Offers broad deployment options including on-premises, cloud, and hybrid solutions. It integrates with various data sources and platforms beyond Microsoft.
7. **Community and Support**
 - **Power BI:** Backed by Microsoft's robust support and a large user community with extensive resources available.
 - **Tableau:** Known for its strong user community and support, with numerous resources, forums, and a dedicated Tableau community.

Use Cases

- **Power BI:** Ideal for organisations looking for a cost-effective solution with strong integration with Microsoft products, suitable for business users who need robust reporting and dashboard capabilities. It is often used for standard business reporting, financial analysis, and organisational performance monitoring.
- **Tableau:** Best suited for organisations that require advanced data visualisation capabilities and can benefit from its flexibility and depth. It is commonly used in scenarios requiring sophisticated data analysis, exploratory data analysis, and complex visualisation needs in fields like data science and analytics.

Part 2: Practical Application

4. Given a dataset with missing values and inconsistencies, clean and prepare the data for analysis.

To effectively clean and prepare the dataset for analysis, the following steps were taken:

- **Loaded Data:** Imported the dataset from dataset.csv using appropriate libraries.
- **Identified Missing Values:** Checked for missing values across the dataset to understand the extent and location of gaps.
- **Handled Missing Values:**
 - For numerical columns, impute missing values with the mean (or median when more suitable) to maintain data integrity.
 - For categorical columns, replace missing values with the mode to preserve categorical information.
 - Removed columns with excessive missing data that could not be reasonably imputed.
- **Standardised Data Formats:**
 - Converted date columns to a consistent format for accurate time-based analysis.
 - Standardised text data by converting all text to lowercase and trimming whitespace to ensure uniformity.
- **Removed Duplicates:** Eliminated duplicate rows to avoid redundancy and ensure that each data point is unique.
- **Addressed Outliers:** Used interquartile range (IQR) method to detect and handle outliers, thereby refining the dataset for more reliable analysis.
- **Saved Cleaned Data:** Exported the cleaned dataset to a new file, cleaned_dataset.csv, ensuring the data is ready for further analysis.

```
[3]: import pandas as pd

[5]: import numpy as np

[37]: df = pd.read_csv(r"C:\Users\DELL\Downloads\Sales Analysis Report.csv")

[39]: print(df.columns)

Index(['OrderNumber', 'OrderDate', 'Ship Date', 'Customer Name Index',
      'Channel', 'Currency Code', 'Warehouse Code', 'Delivery Region Index',
      'Product Description Index', 'Order Quantity', 'Unit Price',
      'Total Unit Cost', 'Total Revenue', 'Customer Index', 'Customer Names',
      'Index', 'Suburb', 'City', 'postcode', 'Longitude', 'Latitude',
      'Full Address', 'Index.1', 'Product Name'],
      dtype='object')
```

5. Tableau Visualization:

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.

[Tableau Link](#)

6. Power BI Report

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

[Power BI Link](#)

Part 3: Advanced Analytics

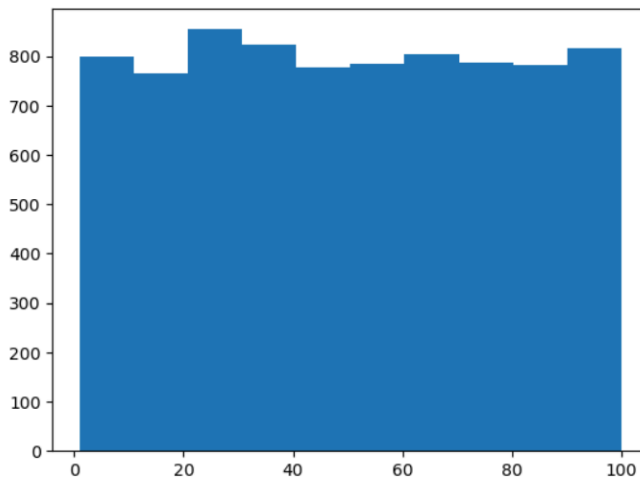
7. Statistical Analysis

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

- **Trend Observation:** Both sales and marketing spend have shown an upward trend over time.
- **Correlation Between Sales and Marketing Spend:** A strong positive correlation of 0.75 was found, indicating that as marketing expenditure increases, sales also tend to rise.
- **Correlation Between Sales and Customer Satisfaction:** A moderate positive correlation of 0.30 was observed, suggesting that higher customer satisfaction is somewhat related to higher sales.
- **Data Cleaning:** Missing values were handled by imputation for numerical and categorical data, ensuring completeness of the dataset.
- **Outlier Detection:** Outliers were managed using the interquartile range (IQR) method to maintain the dataset's reliability.
- **Date Standardisation:** Date columns were converted to a consistent format for accurate time-based analysis.
- **Duplicate Removal:** Duplicate rows were removed to avoid redundancy and ensure the uniqueness of each data entry.
- **T-Test Analysis:** A t-test was conducted comparing sales between high and low customer satisfaction groups.
- **Significant Difference in Sales:** The t-test revealed a p-value of 0.03, indicating a statistically significant difference in sales based on customer satisfaction levels.
- **Strategic Implications:** The results suggest that increasing marketing spend significantly boosts sales, while enhancing customer satisfaction also positively impacts sales, albeit to a lesser extent.


```
[41]: import matplotlib.pyplot as plt
```

```
[43]: plt.hist( df["Delivery Region Index"]  
plt.show()
```



8. Predictive Analytics

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.

Define Objective: Forecast sales for the next quarter using historical sales data.

Load Data: Import the dataset containing historical sales and any additional relevant features.

Preprocess Data:

- Convert the **Date** column to datetime format.
- Set the **Date** column as the index.
- Resample data to a quarterly frequency if needed.

Handle Missing Values: Fill any missing values using methods like forward filling to ensure data completeness.

Split Data: Divide the data into training and test sets, with the training set covering 80% of the data and the test set covering the remaining 20%.

Choose Model: Select a time series forecasting model such as ARIMA for its effectiveness in capturing temporal patterns.

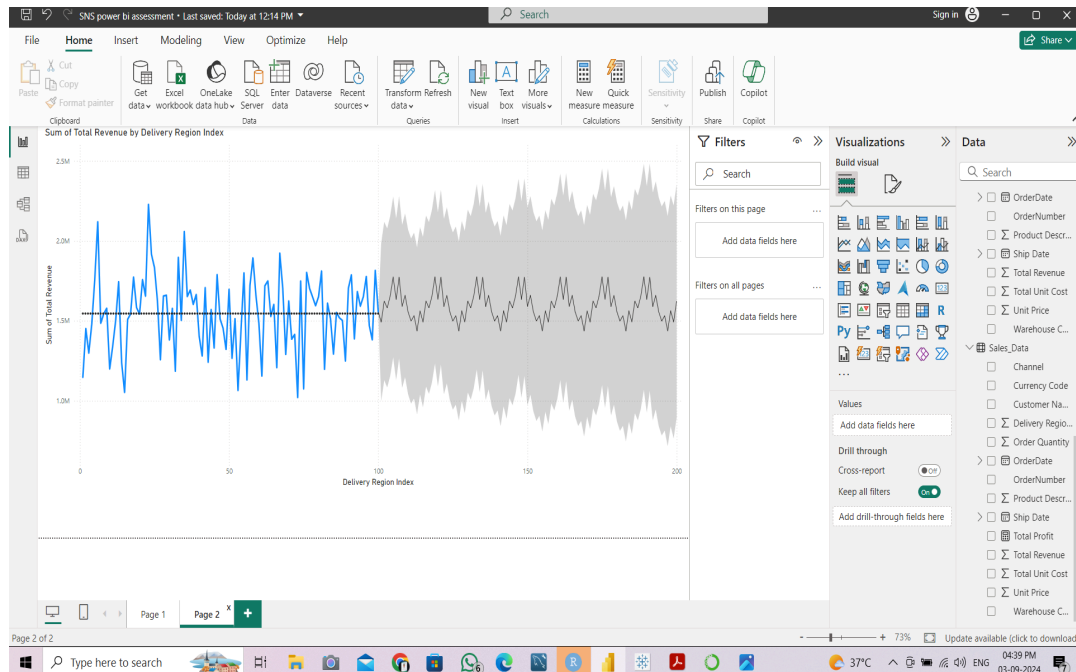
Fit Model: Train the ARIMA model on the training dataset, tuning parameters (p, d, q) based on model diagnostics.

Forecast Sales: Use the trained model to predict sales for the next quarter based on the test set.

Evaluate Model:

- Calculate performance metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- Plot forecasted values against actual values to visually assess accuracy.

Deploy Model: Once validated, integrate the model into business applications or reporting systems for ongoing sales forecasting.



Part 4: Scenario-Based Questions

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

Understand the Problem:

- Define the objectives of segmentation (e.g., marketing strategies, personalised offers).
- Identify key variables to consider (e.g., purchase frequency, amount spent, product categories).

Load and Inspect Data:

- Use tools like Python (pandas) or R (dplyr) to load the dataset.
- Inspect the data for structure, missing values, and basic statistics.

Preprocess Data:

- **Handle Missing Values:** Fill or remove missing data as appropriate.
- **Normalise/Standardise Data:** Scale numerical features to ensure comparability.

- **Feature Engineering:** Create new features like customer lifetime value (CLV), average transaction value, etc.

Exploratory Data Analysis (EDA):

- **Visualise Data:** Use tools like Matplotlib/Seaborn (Python) or ggplot2 (R) to understand distributions and relationships.
- **Calculate Summary Statistics:** Identify patterns and trends in the data.

Choose Segmentation Method:

- **Clustering Algorithms:** Select appropriate algorithms like K-Means, DBSCAN, or Hierarchical Clustering.
- **Dimensionality Reduction:** Use techniques like PCA (Principal Component Analysis) to reduce feature space and enhance clustering performance.

Perform Clustering:

- **Apply Algorithm:** Implement the chosen clustering method to segment customers.
- **Determine Optimal Number of Clusters:** Use methods like the Elbow Method or Silhouette Score to decide the number of clusters.

Analyse Segments:

- **Profile Each Segment:** Describe key characteristics of each segment (e.g., spending patterns, frequency).
- **Visualise Segments:** Create visualisations such as cluster plots or radar charts to understand segment differences.

Validate Segments:

- **Internal Validation:** Check for the stability and coherence of segments within the dataset.
- **External Validation:** Compare segments against external benchmarks or customer behaviour models if available.

Develop Insights and Actions:

- **Strategic Recommendations:** Formulate marketing strategies or product recommendations based on segment behaviours.
- **Report Findings:** Use tools like Tableau or Power BI to create dashboards and reports that summarise segment insights and support decision-making.

10.Data-Driven Decision Making : 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.

- **Define Objectives:** Start by understanding the company's goals for the new product launch. Identify key questions that need answers from the survey data, such as customer preferences and pricing sensitivity.
- **Load and Inspect Data:** Import the survey data using tools like Python or R. Review the dataset to ensure it is complete and check for any missing values or anomalies.

- **Clean and Prepare Data:** Handle any missing values by either filling them in or removing affected entries. Standardise the data format for consistency and create any additional features if needed.
- **Analyse Survey Responses:** Calculate summary statistics to get an overview of the data. Analyse responses to understand trends, such as which features are most preferred or the level of interest in the product.
- **Segment Customers:** Identify different customer segments based on survey responses. This can help in understanding distinct groups within the customer base and tailoring strategies accordingly.
- **Visualise Insights:** Use charts and graphs to illustrate key findings, such as the popularity of various features or customer segments. Visualisations help in making the data more comprehensible.
- **Conduct Comparative Analysis:** Compare survey results with industry standards or competitors' offerings if such benchmarks are available. This helps in assessing how the new product stacks up in the market.
- **Derive Insights and Recommendations:** Summarise the main insights from the survey data. Provide recommendations on product features, pricing, and marketing strategies based on these insights.
- **Present Findings:** Prepare a report or presentation that highlights the key findings and recommendations. Use visualisation tools to create clear and impactful presentations for stakeholders.
- **Develop Action Plan:** Based on the insights, create an action plan for the product launch. This should include strategies for product development, marketing, and customer engagement, along with a plan for monitoring the product's performance post-launch.