Report for Data Analyst role at SNS Group

PART 1

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

Ans- Data visualization helps to tell stories by transforming complex datasets into visual formats that are easier to understand, interpret, and communicate. A good visualization tells a story, removing the noise from data and highlighting useful information.

Key Principles are:

- Diagram First
- Best-suited plots to show a particular trend
- Use of colors
- Simple Visuals, Detailed Captions
- Clarity

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

Ans- There are seven types of Tableau components.

- Tableau Desktop
- Tableau Server
- Tableau Public
- Tableau Reader
- Tableau Online
- Tableau Mobile
- Tableau Data prep

steps for creating basic dashboards in Tableau:

a. Connect to Data: Open Tableau Desktop and connect to your data source. You can choose from various options like Excel, SQL databases, Google Sheets, etc.

b. Prepare Your Data: Clean and transform your data as needed using Tableau's data preparation tools.

c. Creating Data Visualization Sheets: Drag and drop fields onto the Rows and Columns shelves to build charts.

d. Create a Dashboard: Drag and drop the worksheets you created onto the dashboard canvas.

e. Add Interactivity: Add interactive elements such as filters, objects, etc.

f. Publish and Share: Once you are satisfied with your dashboard, you can publish it to Tableau Server, Tableau Online, or Tableau Public.

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

Ans. The main features of PowerBI are:

- **Power Query:** This data mashup and transformation tool integrates data from multiple sources.
- **Power Pivot:** This is a memory tabular data modeling tool.
- **Power View:** This tool lets users create and interact with their own data visualizations.
- **Power Map:** Power Map is a visualization tool for 3D geospatial data that bolsters predictive analytics.
- **Power BI Q&A:** This natural language query engine can take user inputs and produce data visualization outputs.
- **Power BI Embedded:** This tool lets users take customer-facing reports, insights, visuals and other items from Power BI and embed them into other apps or websites.

  PowerBI vs Tableau

- **Performance and Scalability**:

  Power BI: Generally performs well with moderate-sized datasets. It integrates closely with Azure for handling larger datasets and complex queries.

  Tableau: Handles large datasets and complex visualizations very efficiently, often praised for its performance in high-demand scenarios.

- **Pricing**:

  Power BI: Typically considered more cost-effective, with a lower entry price point, especially for organizations already using Microsoft products..

  Tableau: Often seen as more expensive, with pricing that varies based on user roles and features

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 behaviors? Describe thesteps and tools you would use. Evaluation Criteria: Problem-solving approach, understanding of segmentation techniques, choice of tools.

Ans- Identifying key customer segments and understanding their behaviors from a large dataset is a crucial task in data analysis and can significantly impact business strategy. Here's how I would approach the task:

Tools required: python ,powerBi, SQL, Tableau

a. Understanding and Cleaning the Dataset: Review the dataset to understand the available variables such as transaction details (date, amount, product), customer demographics (age, gender, location), and behavioral data (purchase frequency, channels used).

b. Exploratory Data Analysis (EDA): Calculate basic statistics (mean, median, mode, standard deviation) to understand the distribution of key variables like transaction amounts and frequencies.

c. Visualization: Use visual tools like histograms, box plots, and scatter plots to uncover patterns.

d. Feature Engineering:

- **Derive New Metrics**: Create new features such as:
    - **Recency, Frequency, Monetary (RFM)**: Recency of the last purchase, frequency of purchases, and monetary value spent.
    - **Customer Lifetime Value (CLV)**: Estimate the total revenue a business can expect from a customer over the course of their relationship.
    - **Churn Indicators**: Identify customers who haven't transacted in a while as potential churn risks.
- **Normalization/Standardization**: Scale the data to ensure that features like transaction amount and frequency are comparable when clustering.

e. Segmentation Techniques:

- **K-Means Clustering**: Perform clustering to group customers based on similarity across key features (e.g., RFM values). K-Means is widely used for its simplicity and effectiveness in segmentation.
- **Hierarchical Clustering**: Useful for understanding the potential number of clusters by visualizing a dendrogram.
- **DBSCAN**: If the dataset contains noise and outliers, DBSCAN can be useful as it is density-based and doesn't require specifying the number of clusters.
- **PCA (Principal Component Analysis)**: Use PCA to reduce the dimensionality of the dataset, which helps in visualizing and understanding the clusters formed.

f.Implementation and Monitoring:

- **Actionable Insights**: Translate segment characteristics into actionable business strategies. For example, target high-value customers with loyalty programs or re-engage churn-prone segments with special offers.

Conclusion: Data Storytelling- Clearly communicate findings to stakeholders through visualizations, reports, and presentations that highlight key insights and recommended actions.

10. 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. Evaluation Criteria: Approach to data analysis, use of visualization tools, and ability to derive insights.

Ans- When a company is preparing to launch a new product, survey data on customer preferences can be invaluable for making informed decisions. Here's a structured approach to using this data effectively:

a. Understand the Objectives:

- Clarify Business Goals: Understand what the company aims to achieve with the new product launch.
- Survey Design Review: Ensure that the survey data is robust, covering key areas like customer demographics, preferences, price sensitivity, and potential use cases for the product.

b. Data Preparation and EDA

c. Customer Segmentation:

- **Cluster Analysis**: Use clustering techniques (e.g., K-Means, hierarchical clustering) to group customers based on their preferences, willingness to pay, or other survey responses.
- **Persona Development**: Create customer personas representing key segments, such as "Budget-Conscious Buyers," "Premium Seekers," or "Tech Enthusiasts." This helps in tailoring the product offering and marketing strategies.

d. Price Sensitivity Analysis: Analyze survey data to understand how price impacts customer interest. Tools like the Van Westendorp Price Sensitivity Meter can help determine optimal pricing.

e. Competitor Benchmarking: Compare customer preferences with existing products in the market. Identify gaps or opportunities where the new product can offer a unique value proposition.

f. Predictive Modeling:

- **Logistic Regression**: Use logistic regression to predict the likelihood of purchase based on different customer characteristics and preferences.
- **Decision Trees**: Employ decision trees to identify key decision points that influence whether customers would buy the product.
- **Market Simulation**: Use predictive models to simulate market scenarios, such as how changes in pricing or features might impact overall demand.

g. Risk Assessment: Identify potential risks (e.g., low adoption, high price sensitivity) and develop strategies to mitigate them.

h. Conclusion:

- **Data Storytelling**: Present the findings to stakeholders using a combination of visualizations and narratives. Clearly communicate the key insights, the rationale behind recommendations, and how they can be implemented.
- **Dashboard Creation**: Create a dashboard (using tools like Tableau or Power BI) that allows stakeholders to interact with the data and explore different scenarios.

Note- Parts 2 and 3 are practical questions and their files are attached to the task file.