**DCIT 422**

**ASSIGNMENT 1**

**QUESTION ONE**

**The History of Tableau: A Leading Data Visualization Software**

“Tableau, a prominent data visualization software, has revolutionized the way organizations understand and interpret data. Founded in 2003, Tableau's journey from a small startup to a global leader in business intelligence (BI) is a testament to its innovation and adaptability.

**Origins and Foundation (2003-2004)**

Tableau was born out of a Stanford University research project aimed at improving data exploration and visualization. Dr. Pat Hanrahan, a computer graphics professor, and his Ph.D. student, Chris Stolte, developed a breakthrough technology known as VizQL (Visual Query Language). VizQL translated data queries into visual representations, making it easier for users to interact with their data. Recognizing the commercial potential of their invention, Hanrahan and Stolte partnered with Christian Chabot, a business executive, to found Tableau Software in January 2003.

**Early Development and Growth (2004-2010)**

Tableau's initial product, Tableau Desktop, was released in 2004. It allowed users to connect to various data sources and create interactive visualizations without needing advanced technical skills. This ease- of- use distinguished Tableau from other BI tools, which often required extensive programming knowledge. The company's focus on user-friendly design and powerful analytics capabilities resonated with businesses seeking to leverage data for decision-making. Tableau rapidly gained traction, attracting a loyal customer base across diverse industries. In 2008, Tableau introduced Tableau Server, enabling organizations to share and collaborate on data visualizations over the web. This expansion into server-based solutions marked a significant milestone, positioning Tableau as a comprehensive BI platform.

**Expansion and Innovation (2010-2018)**

The 2010s were a period of significant growth and innovation for Tableau. The company went public in May 2013, raising $254 million in its initial public offering (IPO). This infusion of capital fueled further development and expansion efforts. Tableau continued to enhance its product offerings with features like Tableau Public (2010), a free platform for sharing visualizations online, and Tableau Online (2013), a cloud-based version of Tableau Server. These innovations made Tableau accessible to a broader audience, including small businesses and individual users. In 2015, Tableau introduced Tableau Prep, a data preparation tool designed to simplify the process of cleaning and organizing data. This addition addressed a critical need in the data analytics workflow, making it easier for users to prepare their data for analysis.

**Acquisition by Salesforce and Future Prospects (2019-Present)**

In June 2019, Salesforce, a leading customer relationship management (CRM) company, acquired Tableau for $15.7 billion. This acquisition integrated Tableau's powerful data visualization capabilities with Salesforce's extensive CRM platform, creating new opportunities for data-driven insights and enhanced customer experiences. Since the acquisition, Tableau has continued to innovate and expand its offerings. The company has focused on integrating artificial intelligence (AI) and machine learning (ML) capabilities into its platform, enabling users to derive deeper insights from their data. Features like Explain Data, which automatically generates explanations for data points, and Tableau CRM (formerly Einstein Analytics), which combines Tableau's visualization tools with Salesforce's AI capabilities, exemplify this trend.

**QUESTION TWO**

The following below are four commonly used techniques of data visualization:

* **Bar Charts**

Description: Bar charts represent data with rectangular bars, where the length or height of each bar is proportional to the value it represents. They are commonly used to compare different categories or track changes over time.

Use Case: Comparing the sales performance of different products in each period.

* **Line Charts**

Description: Line charts display information as a series of data points connected by straight

lines. They are useful for showing trends over time and are ideal for continuous data.

Use Case: Visualizing the monthly revenue growth of a company over several years.

* **Scatter Plots**

Description: Scatter plots use dots to represent the values of two different variables. The position of each dot on the horizontal and vertical axes indicates the values for an individual data point. Scatter plots are useful for identifying relationships or correlations between variables.

Use Case: Analyzing the relationship between advertising spend and sales revenue.

* **Pie Charts**

Description: Pie charts are circular charts divided into sectors, each representing a proportion of the whole. They are useful for showing the composition of a dataset or the relative proportions of different categories.

Use Case: Displaying the market share of different companies within an industry.

**QUESTION THREE**

Data cleaning is a crucial step in the data visualization process because it ensures the accuracy, reliability, and clarity of the insights derived from the data. Here are several reasons why data cleaning is important:

* **Accuracy and Reliability:**

Error Reduction: Raw data often contains errors such as duplicates, inconsistencies, and missing values. Cleaning the data helps eliminate these errors, ensuring that the visualizations are based on accurate and reliable information.

Consistent Data: Data cleaning standardizes the data, making it consistent across all fields and records. This consistency is essential for accurate analysis and meaningful visual comparisons.

* **Improved Insights:**

Clear Patterns and Trends: Clean data allows for the clear identification of patterns, trends, and relationships within the data. Dirty or inconsistent data can obscure these insights, leading to incorrect conclusions.

Enhanced Decision-Making: Reliable visualizations based on clean data provide better support for decision-making processes. Organizations can make informed decisions with confidence, knowing that their data is accurate and trustworthy.

* **Efficiency:**

Streamlined Analysis: Clean data reduces the time and effort required for data analysis and visualization. Analysts can focus on interpreting the data and generating insights rather than spending excessive time on data preparation.

Automation Readiness: Clean data is essential for automating data analysis and visualization processes. Automated systems rely on consistent and error-free data to function correctly.

* **Effective Communication:**

Clarity and Comprehensibility: Visualizations are meant to communicate information effectively. Clean data ensures that the visual representations are clear, easy to understand, and free from misleading artifacts.

Credibility: Presenting clean, accurate data enhances the credibility of the visualizations and the analysts or organizations that produce them. Stakeholders are more likely to trust and act upon the presented insights.

* **Compliance and Reporting:**

Regulatory Requirements: Many industries have strict regulations regarding data accuracy and reporting. Cleaning data ensures compliance with these regulations, avoiding potential legal and financial penalties.

Accurate Reporting: Clean data is essential for generating accurate reports. Reports based on flawed data can lead to incorrect assessments and misinformed strategies.

**QUESTION FOUR**

* Nominal Data: Nominal data is categorical data without any intrinsic order or ranking among the categories. The categories are simply labels or names, and there is no logical sequence. Example: Colors (red, blue, green), types of fruits (apple, banana, cherry).
* Ordinal Data: Ordinal data is categorical data with a meaningful order or ranking among the categories. The categories can be arranged in a logical sequence, but the intervals between them are not necessarily equal or meaningful. Example: Survey ratings (poor, fair, good, excellent), education levels (high school, bachelor's, master's, doctorate).

**QUESTION FIVE**

The following below are python libraries useful for visualization.

* Matplotlib
* Seaborn
* Plotly
* Altair”