**2nd Semester B. Tech**

**DSE 1271 Data Visualization [1 0 3 2]**

**Week 10 & 11 Assignment: Marks(10)**

**Task:** For the dataset given (SET 4), answer the questions with a report that contains a story, visuals, and data summaries to provide actionable insights into the data.

**Dataset: Transaction Data**

For the dataset transaction\_data.csv (Set 4), answer the following questions with a report that includes insights, visuals, and data summaries to provide actionable insights.

**Dataset Overview: (Set 4)**

The dataset contains details about various financial transactions, including transaction types, amounts, statuses, fraud flags, geolocation, and device usage.

The dataset transaction\_data.csv contains the following attributes:

1. Transaction ID – Unique identifier for each transaction.
2. Sender Account ID – The account ID of the sender.
3. Receiver Account ID – The account ID of the receiver.
4. Transaction Amount – The amount of money transferred in the transaction.
5. Transaction Type – The type of transaction (e.g., Deposit, Withdrawal, Transfer).
6. Timestamp – The date and time of the transaction.
7. Transaction Status – The status of the transaction (e.g., Success, Failed).
8. Fraud Flag – Indicates whether the transaction was flagged as fraudulent (True/False).
9. Geolocation (Latitude/Longitude) – The location from which the transaction was initiated.
10. Device Used – The device used for the transaction (e.g., Desktop, Mobile).

**Questions**

1. Perform Data Exploration on the dataset: **(2M)**

2. Perform the necessary cleaning, in terms of missing and noisy data. Summarize the data before and after cleaning. **(2M)**

**3.** Use Python and libraries like Pandas, Matplotlib and Seaborn to create insightful tabulations and visualization for the following:

**(4M for analysis + 2M for report and notebook submission = 6M)**

1. **Distribution of Attributes**  
   **Example Insights.** What is the overall distribution of Transaction Amount? Are there extreme values or outliers in Transaction Amount? What is the proportion of different Transaction Types? What percentage of transactions are flagged as fraud?
2. **Correlation Among Attributes**

Example Insights: How does Transaction Amount correlate with fraud? Does Device Used affect transaction success rates? Is there a relationship between Geolocation and fraud occurrences?

1. **Analyzing Feature Interactions**

Example Insights: Do higher Transaction Amounts have a higher fraud rate? How does Transaction Type impact Transaction Status? Are mobile users more likely to experience failed transactions compared to desktop users?

**Submission on LMS**

* Python notebook to showcase the process of exploring, cleaning, and analysis of the dataset. From the data summaries, charts/plots and analysis of the dataset, gather the insights and write as markdown (in text) in the notebook.
* A report (in pdf) containing a narrative considering the audience is the Sales manager. You are presenting the findings and insights from the data analysis of the Northwind dataset.