Key Insights to Big Data Project

Name	Sec	B.N	Code
Yousef Osama Mohamed	2	32	9211386
Nesma Abdelkader	2	28	9211292
Sara Gamal Gerges	1	20	9210455
Eman Ibraheem	1	14	9210265

Fraud Detection System

Problem Idea:

This project aims to enhance financial services across three core areas: fraud detection, customer analytics, and financial forecasting. It aims to build intelligent systems for real-time fraud prevention, personalized customer insights, and predictive financial planning—empowering businesses to improve security, optimize customer engagement, and make data-driven financial decisions.

Data Set:

- **Main Dataset:** Transactions fraud dataset
- **Size:** 8M rows, 39 columns (about +20 useful features after excluding columns like card_id, transaction_id, and user_id, etc), 1.42 GB.

Questions we will answer:

✓ What patterns indicate fraudulent transactions?
✓ How can behavior segment customers?
✓ What will the user's credit score be based on their transactions and data?

EDA (Exploratory Data Analysis):

1. Merge datasets using relational keys (e.g., user_id, card_id).
2. Handle missing values and clean anomalous entries.
3. Visualize transaction distribution across time, location, and merchant.
4. Analyze class imbalance and identify correlation patterns.
5. Explore user-level and merchant-level behavioral trends.

Descriptive Analysis Methods:

1. **Behavioral Profiling:**
   - ☐ Summary statistics of typical user activity
2. **Association Rules (Data mining):**
   - ☐ specific merchant → high amount transaction
   - ☐ online transaction, specific city → fraud transactions

Predictive Analysis Methods:

a. XGBoost

b. Naïve Bayes Classifier [Map reduced]

c. Logistic Regression & linear regression [Map reduced]

d. Random Forest [Map reduced]

e. KNN [Map reduced]

Key Insights

Transaction Insights

1. **Analysis of Common Transaction Types**
   • **Insight**: Understanding the most frequent transaction types provides valuable insights into customer preferences and behavior.
2. **Association of Transaction Features**
   • **Insight**: Discovering relationships between transaction attributes helps uncover meaningful patterns that can improve decision-making. For example, certain combinations of transaction features (e.g., high-value purchases at specific times or locations).
3. **Time-Based Transaction Behavior Analysis**
   • **Insight**: Analyzing transaction patterns at specific times of the day, week, or year.

4. **Transaction Error Rates**
   • **Insight**: Monitoring transaction errors (e.g., failed transactions, system timeouts, or incorrect processing) can help identify areas for improvement in the transaction process.

User Insights

5. **Spikes in Transaction Frequency per User**
   • **Insight**: A sudden increase in transaction frequency for a specific user may indicate a potential fraud attempt.
6. **Unusual Transaction Amounts per User**
   • **Insight**: Transactions involving significantly higher amounts than usual could be a red flag for fraud.
7. **Uncommon Merchant Interaction Patterns**
   • **Insight**: Frequent or large transactions with unfamiliar or rare merchants might indicate fraudulent activity. Users making purchases at merchants they don’t typically interact with, especially in large volumes, could be flagged for review.

8. **Customer Segmentation Based on Transaction Behavior**
   • **Insight**: Grouping customers by similar spending patterns helps identify trends and predict future behavior.
9. **Predicting Customer Credit Scores**
   • **Insight**: Predicting customer credit scores based on transaction history and other behaviors can help assess financial stability and risk.

Fraud Detection Insights

10. **Fraud Detection in Transactions**
    • **Insight**: Using machine learning models such as logistic regression and XGBoost can enable real-time detection of fraudulent transactions.