Koneru Lakshmaiah Education Foundation
(Category -1, Deemed to be University estd. u/s. 3 of the UGC Act, 1956)
Accredited by **NAAC** as '**A++**' ❖Approved by AICTE ❖ ISO 9001-2015 Certified
Campus: Green Fields, Vaddeswaram - 522 302, Guntur District, Andhra Pradesh, INDIA.
Phone No. 08645 - 350200; www.klef.ac.in; www.klef.edu.in; www.kluniversity.in
Admin Off: 29-36-38, Museum Road, Governorpet, Vijayawada - 520 002. Ph: +91 - 866 - 3500122, 2576129.

Fraud Detection In E-Commerce Transactions

A Project Report

Submitted in the partial fulfilment of the requirements for the

Course Title: Cloud Based Data Analytics Speciality

Course code: 22SDCS08A

Submitted by

Name	Id Number
G. Rishitha	2200033117
M. Varun Chandra Reddy	2200033001
K. Rama Sai	2200031487

UNDER THE GUIDANCE OF

Dr. P. Anusha **Associate Professor**

Department of Computer Science and Engineering

KL UNIVERSITY

Green fields, Vaddeswaram – 522 302

Guntur Dt., AP, India

Certification

This is to certify that the Project Report entitled **"Fraud Detection In E-Commerce Transactions"** is being submitted by **G. Rishitha, M. Varun Chandra Reddy, K. Rama Sai, bearing Registered Number 2200033117, 2200033001, 2200031487** in partial fulfilment for the award of **B. Tech III Semester** in **CSE** to the KL University is a record of Bonafide work carried out under our guidance and supervision. The results embodied in this report have not been copied from any other departments/ University/ Institute.

Signature of Supervisor

Dr. P. Anusha

ABSTRACT

This project aims to enhance the detection of fraudulent transactions in e-commerce and banking by leveraging Microsoft Azure cloud services. It collects, processes, and analyses real-time transactional data to identify suspicious patterns that may indicate fraud. Azure Data Lake is used to securely store large volumes of structured and unstructured transaction data. Azure Synapse Analytics facilitates scalable and efficient data processing and querying to detect anomalies. Power BI is used to visualize critical fraud metrics in interactive dashboards, such as abnormal transaction patterns, high-risk user behaviours, and time-based anomalies. These insights empower financial analysts and security teams to respond quickly, reduce false positives, and prevent financial losses. By integrating Azure cloud services with machine learning algorithms, this system provides a dynamic, scalable, and accurate fraud detection solution. It enables organizations to proactively safeguard digital transactions and maintain trust in financial systems.

Keywords: Microsoft Azure, Azure Data Lake, Azure Synapse Analytics, Power BI, fraud detection, e-commerce, banking, real-time monitoring, anomaly detection, machine learning, transaction data, financial security, data visualization, fraud prevention, cloud-based solution, scalable analytics, suspicious activity, digital finance safety.

CONTENTS

CHAPTER 1: INTRODUCTION

The rapid rise of digital transactions in e-commerce and banking has transformed the financial landscape, offering convenience and speed to users worldwide. However, this shift has also intensified concerns around fraudulent activities, with traditional rule-based systems struggling to detect increasingly complex and evolving threats. As digital fraud becomes more sophisticated, there is a growing demand for intelligent, real-time solutions that can adapt quickly to changing patterns.

This project addresses these challenges by developing a cloud-based fraud detection system using Microsoft Azure services. The solution integrates Azure Data Lake for secure and scalable storage of large volumes of transactional data, Azure Synapse Analytics for advanced data processing and anomaly detection, and Power BI for real-time visualization of fraud insights. Together, these tools enable the collection, analysis, and presentation of critical fraud indicators in a dynamic and actionable format.

Key metrics such as transaction anomalies, unusual spending behavior, and high-risk account activities are continuously monitored to support real-time fraud identification. Financial institutions can leverage these insights to reduce false positives, respond to threats proactively, and improve the accuracy of fraud detection. By utilizing Azure's powerful capabilities, this system offers a scalable,

reliable, and intelligent solution that enhances digital transaction security and protects both organizations and consumers from financial loss.

Keywords: Microsoft Azure, Azure Data Lake, Azure Synapse Analytics, Power BI, fraud detection, e-commerce, banking, cloud computing, machine learning, transaction monitoring, anomaly detection, data visualization, financial security, scalable solution, real-time alerts, intelligent fraud prevention, digital finance, secure transactions.

1.1 Problem Statement

In today's digital age, the rise of online transactions in e-commerce and banking has led to a significant increase in fraudulent activities. Traditional fraud detection systems, which rely on static rule-based methods, often fail to identify complex and evolving threats. These limitations result in delayed detection, high false positive rates, and increased financial losses.

The growing complexity of fraud patterns demands a smarter, real-time approach to detection. Vast amounts of transactional data must be analysed quickly and accurately to spot subtle anomalies and alert stakeholders effectively. Current systems often lack scalability, intelligent analytics, and real-time responsiveness.

1.2 Existing System

Existing fraud detection systems in e-commerce primarily rely on manual or rule-based methods to identify suspicious transactions. These systems often use fragmented detection logic, lack real-time analysis capabilities, and provide limited insights through static or periodic reports.

Key limitations include:

- Difficulty in detecting newly evolving or sophisticated fraud patterns.
- Inability to provide real-time transaction monitoring and alerts.
- Lack of intuitive dashboards for analysts and decision-makers.

1.3 Proposed System

The proposed solution leverages Azure's cloud-based services to streamline the storage, processing, and visualization of transactional data in e-commerce. By integrating Azure Data Lake, Azure Synapse Analytics, and Power BI, the system offers:

- Centralized and secure storage of large volumes of transaction data.
- Real-time data processing and fraud detection using Synapse and ML models.
- Interactive Power BI dashboards for visualizing fraud alerts and system performance.

1.4 Literature Review

Several research studies and real-world implementations have explored the use of cloud platforms and machine learning techniques to enhance fraud detection systems. These studies emphasize the significance of scalable data processing, real-time anomaly detection, and intuitive visualization tools in improving fraud detection accuracy and operational efficiency.

Notable contributions include:

- Research on machine learning algorithms such as Random Forest for accurate fraud classification.
- Projects demonstrating real-time analytics using Azure Synapse and other big data tools.
- Case studies highlighting the role of Power BI dashboards in visualizing fraud patterns and transaction trends.

CHAPTER 2: SYSTEM REQUIREMENTS

2.1 Software Requirements

2.1.1 Azure Portal:

- Acts as the central platform to manage Azure services.
- Provides an intuitive interface for deploying and configuring resources.

2.2 Software Requirement Specification

2.2.1 Azure Data Lake Storage

- Stores raw and processed datasets securely and efficiently.
- Offers scalable and cost-effective solutions for handling large data files.

- Enables seamless integration with other Azure services for advanced analytics.

2.2.2. Azure Synapse Analytics

- Orchestrates and automates real-time data processing and transformation.
- Supports running SQL queries and machine learning models for fraud detection.
- Bridges data storage and visualization layers with optimized performance.

2.2.3. Power BI

- Builds dynamic dashboards and real-time visualizations for fraud monitoring.
- Highlights suspicious activity, transaction trends, and alert statistics.
- Enables analysts and stakeholders to make timely, data-driven decisions.

CHAPTER 3: DATASET AND DESCRIPTION

3.1 Dataset Overview

This dataset contains detailed information about e-commerce transactions, including transaction amounts, types, device information, timestamps, and user identifiers. It is used to distinguish between legitimate and fraudulent activities by analysing behavioural patterns in the data. The dataset is structured in CSV format and includes thousands of records, enabling effective training and evaluation of fraud detection models. It serves as the foundation for implementing and testing real-time fraud analytics and classification techniques.

3.2 Data Source

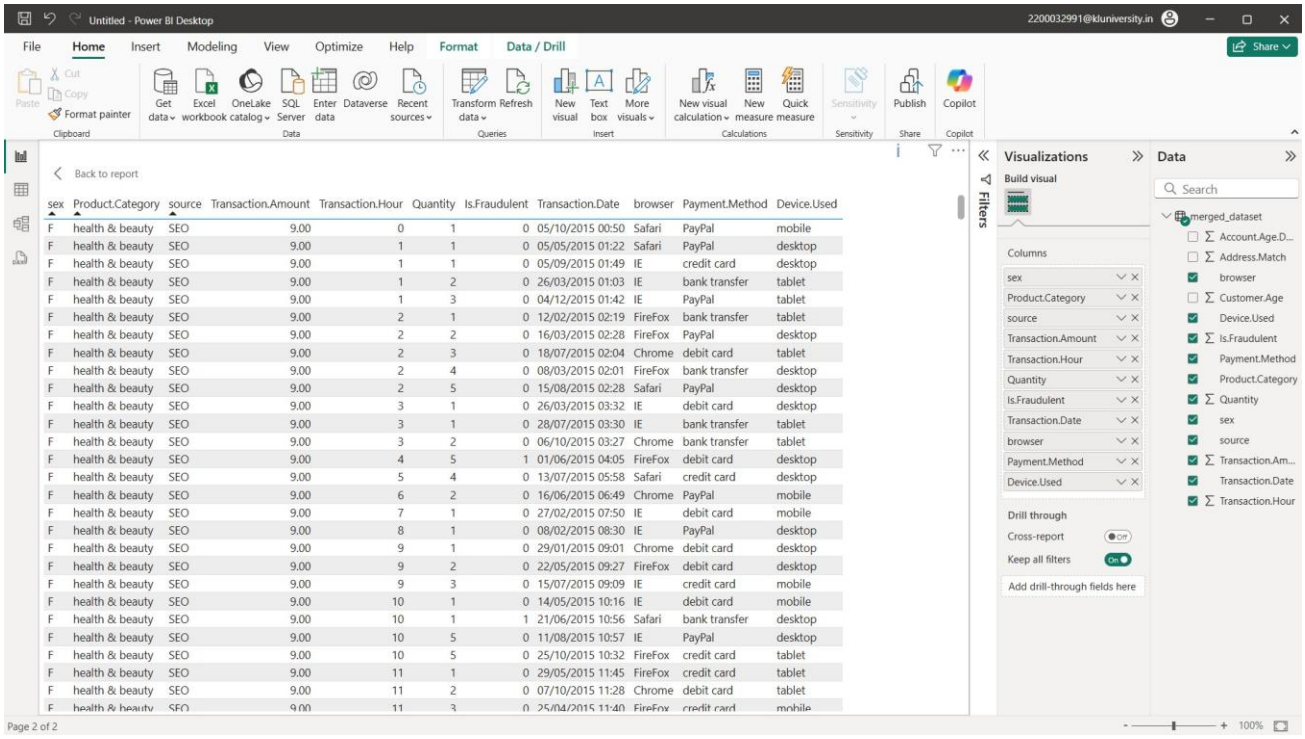
The dataset was sourced from **Kaggle**, containing both genuine and fraudulent transaction records for supervised learning and analytical purposes.

3.3 Attributes and Features

The dataset includes the following columns:

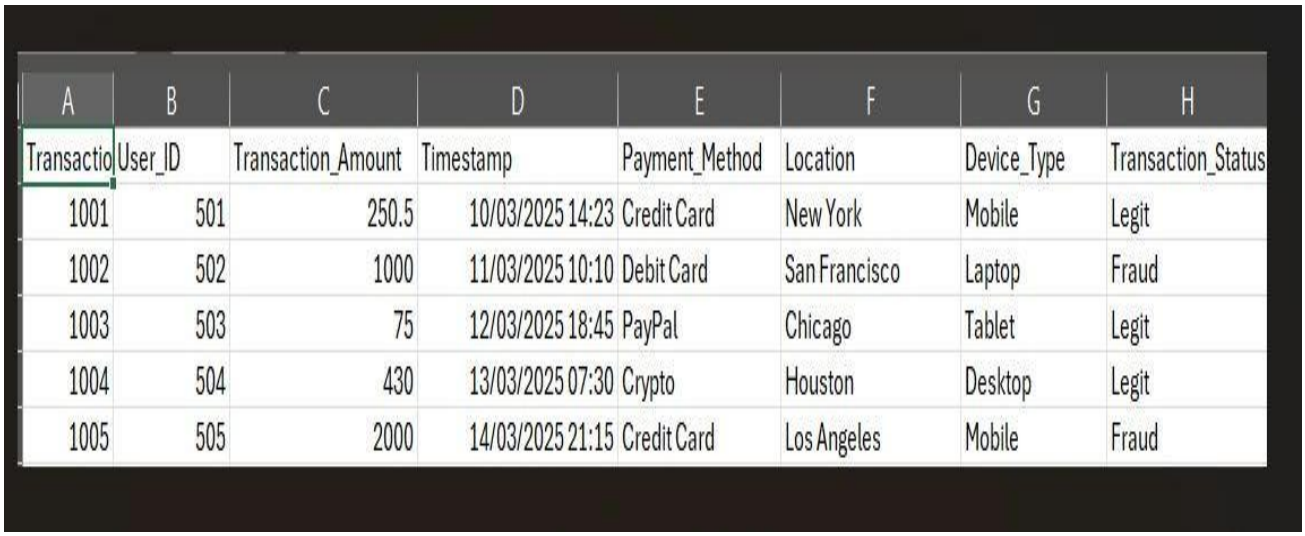
- **TransactionID:** Unique identifier for each transaction.
- **UserID:** Identifier representing the user involved in the transaction.
- **Amount:** Value of the transaction.
- **Timestamp:** Date and time at which the transaction was made.
- **DeviceType:** Device used to perform the transaction (e.g., mobile, desktop).
- **TransactionType:** Type of transaction (e.g., purchase, transfer, login).
- **Location:** Geographical location of the user during the transaction.
- **FraudLabel:** Binary indicator (0 = Legitimate, 1 = Fraudulent).
- **IP_Address**: IP address used at the time of transaction
- **BrowserInfo**: User's browser or app information.

3.4 Dataset Image (E-Commerce.csv)



sex	Product.Category	source	Transaction.Amount	Transaction.Hour	Quantity	Is.Fraudulent	Transaction.Date	browser	Payment.Method	Device.Used
F	health & beauty	SEO	9.00	0	1	0	05/10/2015 00:50	Safari	PayPal	mobile
F	health & beauty	SEO	9.00	1	1	0	05/05/2015 01:22	Safari	PayPal	desktop
F	health & beauty	SEO	9.00	1	1	0	05/09/2015 01:49	IE	credit card	desktop
F	health & beauty	SEO	9.00	1	2	0	26/03/2015 01:03	IE	bank transfer	tablet
F	health & beauty	SEO	9.00	1	3	0	04/12/2015 01:42	IE	PayPal	tablet
F	health & beauty	SEO	9.00	2	1	0	12/02/2015 02:19	FireFox	bank transfer	tablet
F	health & beauty	SEO	9.00	2	2	0	16/03/2015 02:28	FireFox	PayPal	desktop
F	health & beauty	SEO	9.00	2	3	0	18/07/2015 02:04	Chrome	debit card	tablet
F	health & beauty	SEO	9.00	2	4	0	08/03/2015 02:01	FireFox	bank transfer	desktop
F	health & beauty	SEO	9.00	2	5	0	15/08/2015 02:28	Safari	PayPal	desktop
F	health & beauty	SEO	9.00	3	1	0	26/03/2015 03:32	IE	debit card	desktop
F	health & beauty	SEO	9.00	3	1	0	28/07/2015 03:30	IE	bank transfer	tablet
F	health & beauty	SEO	9.00	3	2	0	06/10/2015 03:27	Chrome	bank transfer	tablet
F	health & beauty	SEO	9.00	4	5	1	01/06/2015 04:05	FireFox	debit card	desktop
F	health & beauty	SEO	9.00	5	4	0	13/07/2015 05:58	Safari	credit card	desktop
F	health & beauty	SEO	9.00	6	2	0	16/06/2015 06:49	Chrome	PayPal	mobile
F	health & beauty	SEO	9.00	7	1	0	27/02/2015 07:50	IE	debit card	mobile
F	health & beauty	SEO	9.00	8	1	0	08/02/2015 08:30	IE	PayPal	desktop
F	health & beauty	SEO	9.00	9	1	0	29/01/2015 09:01	Chrome	debit card	desktop
F	health & beauty	SEO	9.00	9	2	0	22/05/2015 09:27	FireFox	debit card	desktop
F	health & beauty	SEO	9.00	9	3	0	15/07/2015 09:09	IE	credit card	mobile
F	health & beauty	SEO	9.00	10	1	0	14/05/2015 10:16	IE	debit card	mobile
F	health & beauty	SEO	9.00	10	1	1	21/06/2015 10:56	Safari	bank transfer	desktop
F	health & beauty	SEO	9.00	10	5	0	11/08/2015 10:57	IE	PayPal	desktop
F	health & beauty	SEO	9.00	10	5	0	25/10/2015 10:32	FireFox	credit card	tablet
F	health & beauty	SEO	9.00	11	1	0	29/05/2015 11:45	FireFox	credit card	tablet
F	health & beauty	SEO	9.00	11	2	0	07/10/2015 11:28	Chrome	debit card	tablet
F	health & beauty	SEO	9.00	11	3	0	25/04/2015 11:40	FireFox	credit card	mobile

Fig 3.4: Dataset Image (E-Commerce.csv)

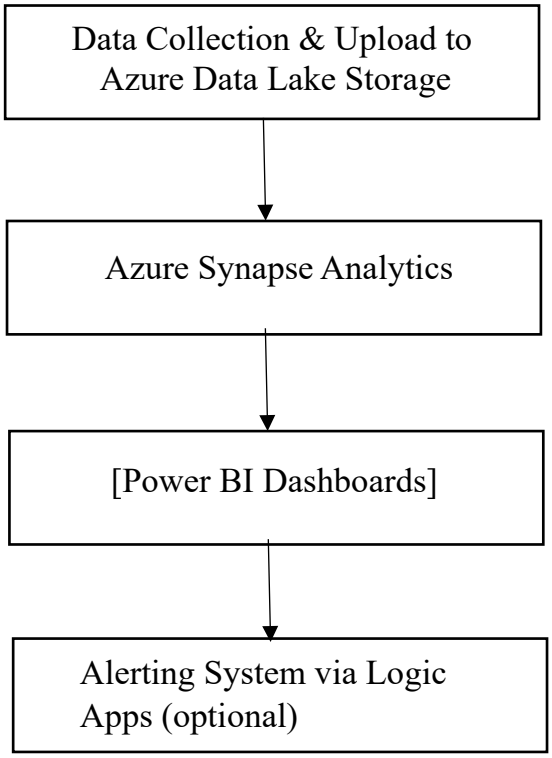


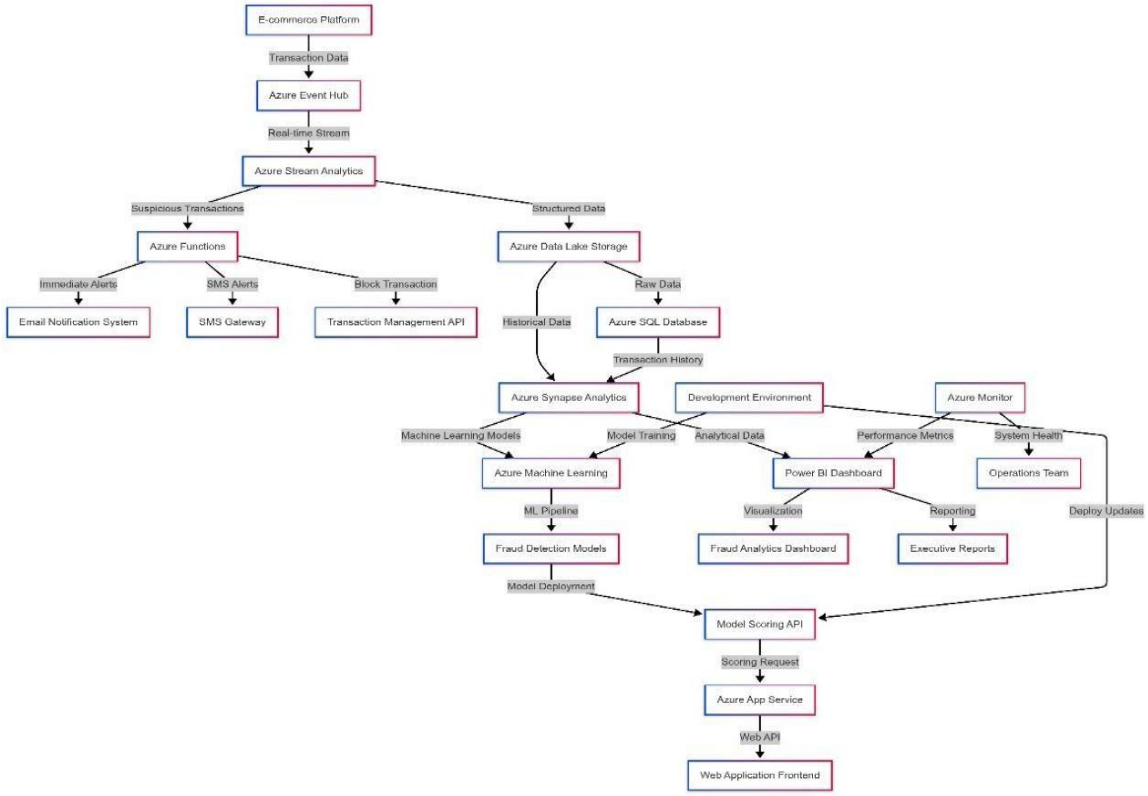
A	B	C	D	E	F	G	H
Transactio	User_ID	Transaction_Amount	Timestamp	Payment_Method	Location	Device_Type	Transaction_Status
1001	501	250.5	10/03/2025 14:23	Credit Card	New York	Mobile	Legit
1002	502	1000	11/03/2025 10:10	Debit Card	San Francisco	Laptop	Fraud
1003	503	75	12/03/2025 18:45	PayPal	Chicago	Tablet	Legit
1004	504	430	13/03/2025 07:30	Crypto	Houston	Desktop	Legit
1005	505	2000	14/03/2025 21:15	Credit Card	Los Angeles	Mobile	Fraud

Fig 3.4: Dataset Image (E-Commerce.csv)

CHAPTER 4: SYSTEM ARCHITECTURE

4.1 Architectural Diagram





4.2 Workflow Overview

4.2.1 Data Collection and Upload to Azure Data Lake

- CSV files containing **transactional data**, including both legitimate and fraudulent transactions, are uploaded to containers in **Azure Data Lake**.
- **Azure Data Lake** provides scalable and secure storage for both structured and unstructured transaction data.
- Containers are organized logically, enabling easy access and management of the dataset for further analysis.

4.2.2 Data Processing with Azure Synapse Analytics

- **Azure Synapse Analytics** extracts raw transaction data from **Azure Data Lake** and performs ETL (Extract, Transform, Load) operations.
- Data is cleaned, transformed (e.g., standardizing formats, handling missing values), and prepared for model training and fraud detection.

  The transformed data is stored back into **Azure Data Lake** for use in real-time fraud detection and visualization.

4.2.3 Visualization with Power BI

- Power BI connects directly to **Azure Synapse Analytics** to import processed transaction data.

- Interactive **dashboards and reports** are created to analyse **fraud detection trends** and **transaction patterns**.
- Insights, such as **fraudulent transaction rates**, **anomaly detection frequency**, and **real-time alerts**, are presented visually for decision-making and quick action.

CHAPTER 5: IMPLEMENTATION

5.1 Setting Up Azure Data Lake

Objective: Create a secure storage system to upload and manage the dataset.

Step 1: Create a Storage Account

- Log in to the Azure Portal.
- Navigate to Storage accounts > Create.
- Fill in the required details:
  - Subscription: Choose your subscription.
  - Resource Group: Create a new one or use an existing group.
  - Storage Account Name: Enter a globally unique name
  - Performance/Redundancy: Choose based on the project (e.g., Standard, LRS for cost-effectiveness).
- Click Review + Create > Create. Wait for deployment.

Step 2: Create a Container

- Open the newly created storage account from the Storage accounts tab.
- Navigate to Containers > + Container.
- Enter a name for the container and set the access level:
    - For private access, choose Private (no anonymous access).
- Click Create.

Step 3: Upload Dataset

- Go to the container you created.
- Click Upload.
- Browse your local machine and select the dataset CSV file.
- Optionally, configure advanced options like metadata or overwrite existing files.
- Click Upload. The dataset is now stored securely in Azure Blob Storage.

5.2 Setting Up Azure Synapse Analytics

Objective: Query and transform transaction data using T-SQL for rule-based fraud detection.

Step 1: Create Synapse Workspace

- Navigate to Synapse Analytics > Create.
    - Provide the required details:
        - Workspace Name: e.g., fraud-synapse.
        - Storage Account: Link the one created earlier.
        - File System: Use the container name created earlier
    - Click Review + Create > Create. Wait for deployment.

Step 2: Query Data Using SQL

- Launch Synapse Studio from the workspace.
- Use the Data Hub to link the storage and explore the CSV file.
- Create an external table using serverless SQL pool.

5.3 Creating Visualizations with Power BI

Objective: Visualize and analyse flagged transactions.

Step 1: Connect Power BI to Azure Data Lake

- Open Power BI Desktop on your local machine.
- Go to Home > Get Data > Azure > Azure Blob Storage.
- Enter the storage account name and authenticate:

- Use the storage account key or SAS token for secure access.

- Browse the container and select the processed fraud detection dataset.

Step 2: Create Data Visualizations

- Load the dataset into Power BI.

- Use the following visuals:

  - Bar Charts: For transaction patterns.

  - Pie Charts: To show fraud detection percentages across transactions.

  - Tables: To list individual transaction details.

- Format visuals for clarity using Power BI's formatting tools.

Step 3: Publish the Dashboard

1. Save the report and click Publish.

2. Log in to the Power BI Service and choose a workspace.

3. Share the report with stakeholders by configuring access permissions.

5.4 Automating the Workflow

Objective: Enable scheduled updates for data processing and reporting.

Step 1: Automate Data Updates in Synapse

- Add an external data refresh process or use Azure Data Factory (optional) to reload updated CSVs into Synapse.

Step 2: Enable Data Refresh in Power BI

- Log in to Power BI Service.
  - Go to the published report > Settings > Datasets.
  - Enable scheduled refresh to pull updated data from Azure Data Storage.

CHAPTER 6: RESULTS

6.1 Output Screenshots

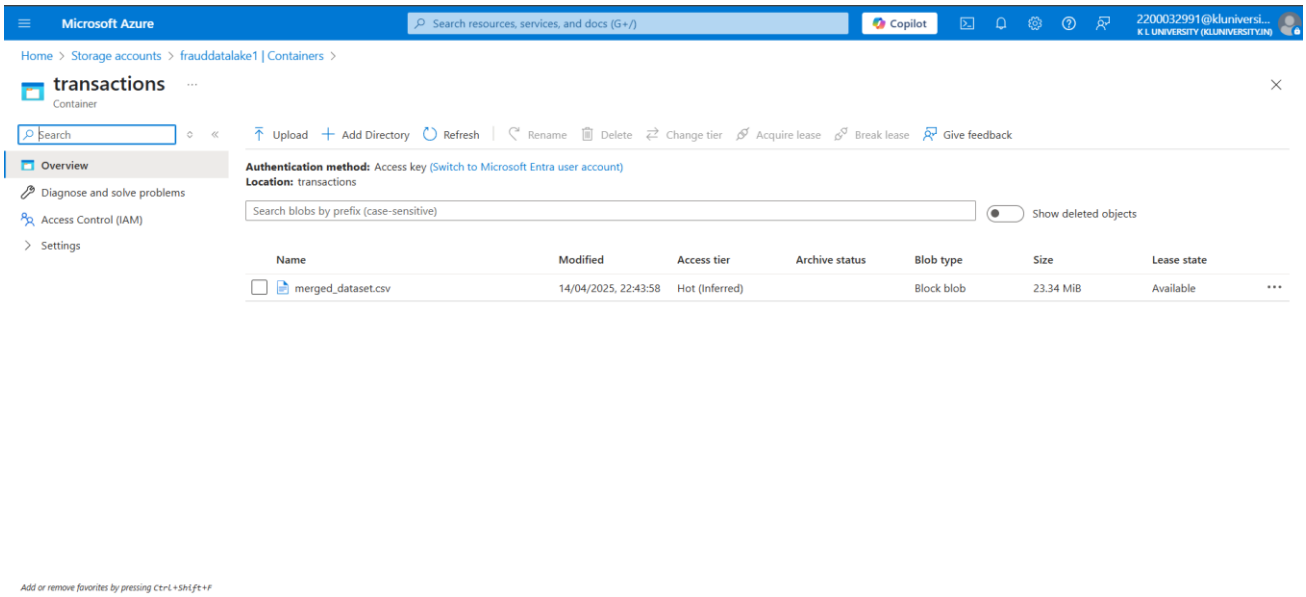


Fig 6.1.1: Uploaded Dataset

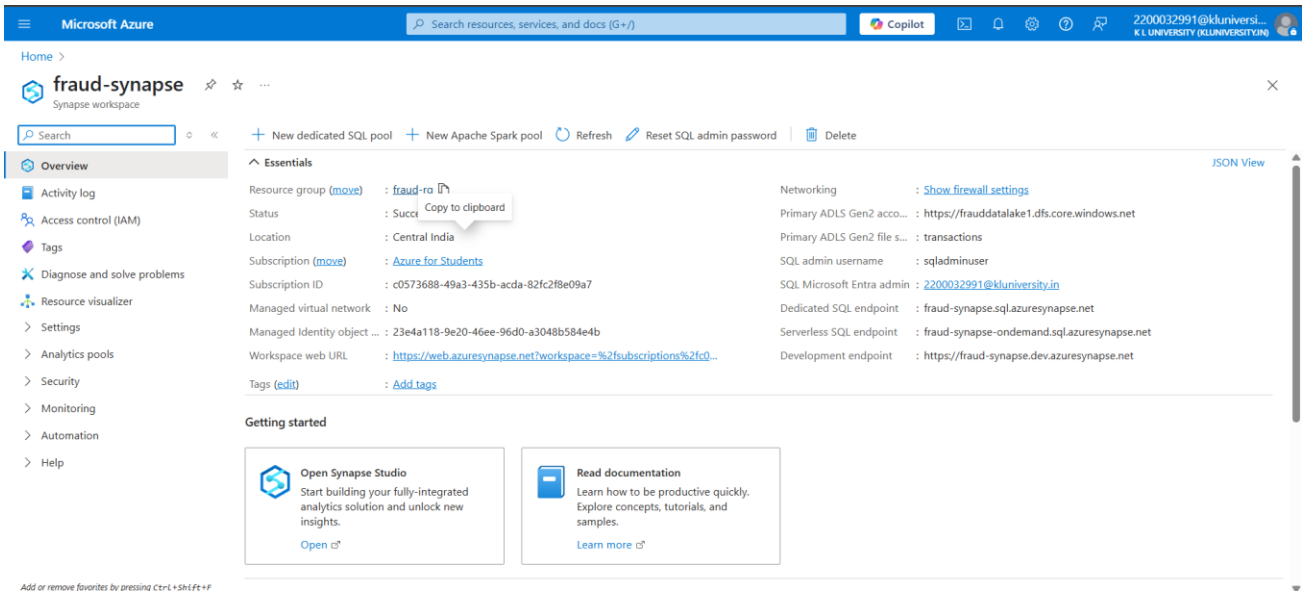


Fig 6.1.2: Creating a Azure Synapse Analysis

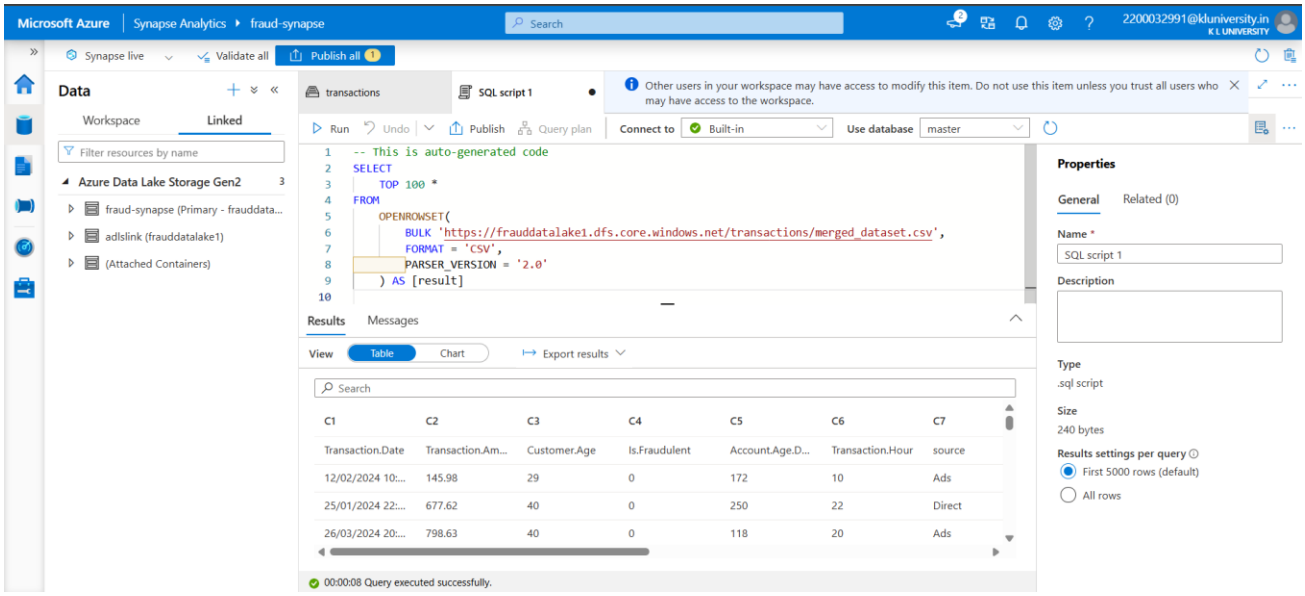


Fig 6.1.3: Integrating dataset into synapse analytics from Data Lake Storage

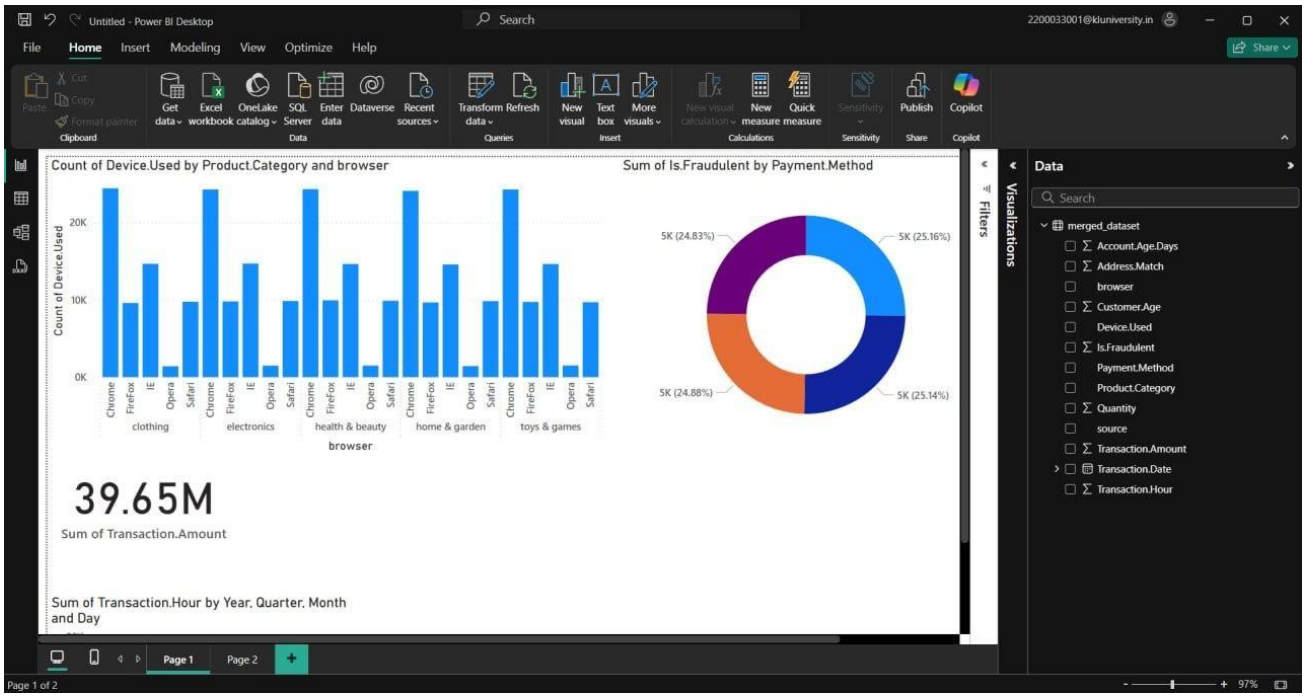


Fig 6.1.4: Transaction Visualization

6.2 Dashboard Analysis

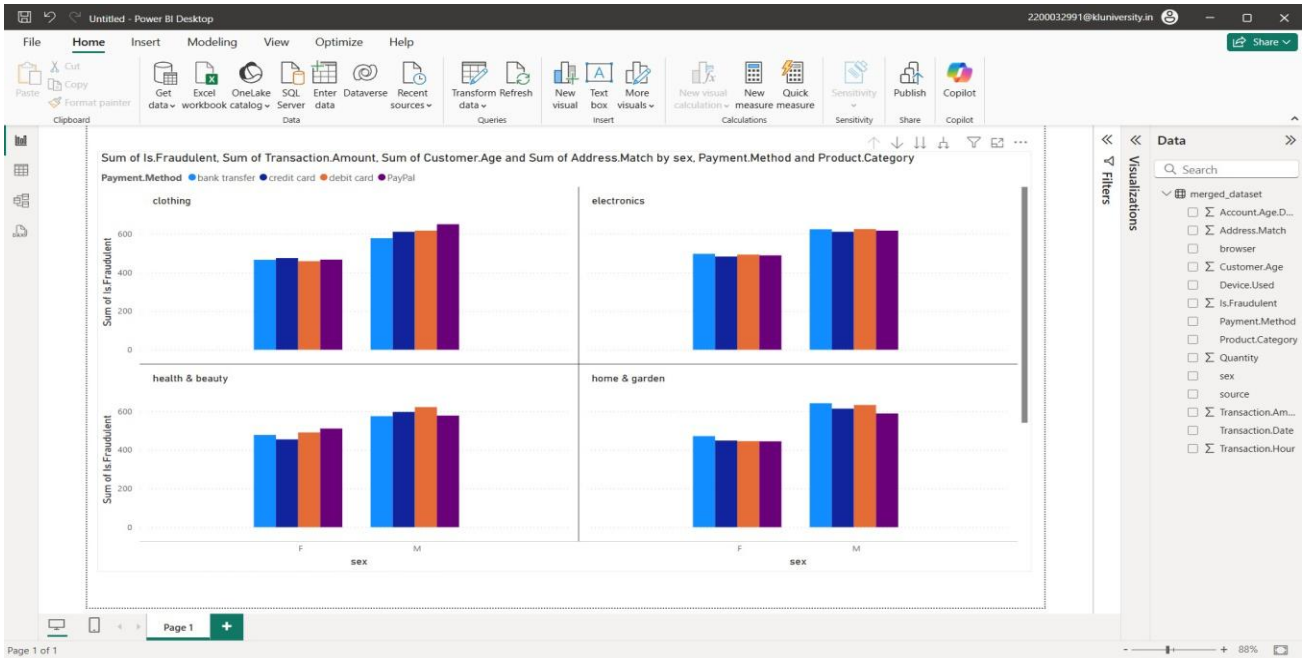


Fig 6.1.5: Dashboard Analysis

Fig 6.1.4: At 25.16%, Credit Card had the highest Sum of Is Fraudulent by Payment. Method, which was slightly higher than the lowest, Cash, at 24.83%. Credit Card accounted for the highest portion of fraudulent transactions at 25.16%, followed closely by PayPal at 25.14%. Cash had the lowest representation in fraud cases at 24.83%. Across all payment methods, the Sum of Is. Fraudulent ranged from 24.83% to 25.16%.

Fig 6.1.5: In terms of device usage by product category and browser, Chrome on electronics showed one of the highest device usage counts, exceeding 20K. Safari in clothing and Internet Explorer in electronics had comparatively lower counts, each under 10K. The total Sum of Transaction, Amount was 39.65 million. Transaction activity by hour is visualized further below, showing trends by year, quarter, month, and day.

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

The project successfully demonstrated a cloud-based fraud detection system for e-commerce transactions. The integration of Azure services provided a robust framework for data storage, processing, real-time analytics, and visualization.

Future Scope:

- Integration of predictive analytics using machine learning models.
- Real-time data monitoring for instant feedback.
- Scalability to include data from multiple e-commerce platforms and broader datasets.

7.1 Challenges Faced

- **Data Integration Complexity**: Combining structured transaction data from various sources into Azure Data Factory involved complex data transformations, including mapping different formats and ensuring consistency for accurate fraud detection.
- **Storage Management**: Ensuring efficient organization of large datasets in **Azure Data Lake Storage** posed challenges in terms of access control and ensuring the smooth handling of transaction data while maintaining security and performance.

- **Performance Optimization**: Training the **Random Forest classifier** and optimizing it to detect fraudulent patterns accurately in large datasets required iterative testing and parameter tuning to achieve high precision.
- **Dashboard Design**: Designing an intuitive yet informative **Power BI dashboard** that could present fraud alerts, transaction trends.
- **Data Security**: Configuring secure access using permissions in Azure Data Lake Storage required thorough testing.

References:

[1] E-learning Dataset Sources: Dataset from Kaggle

[2] Microsoft Azure Documentation: https://learn.microsoft.com/en-us/azure/

[3] Power BI Official Documentation: https://learn.microsoft.com/enus/power-bi/

[4] Azure Data Factory Tutorials: https://learn.microsoft.com/en-us/azure/datafactory/

[5] Blob Storage Quickstart Guide: https://learn.microsoft.com/enus/azure/storage/blobs/

[6] Data Analysis and Visualization Blogs: https://www.dataversity.net/

[7] IEEE Research Paper on Fraud Detection Models: https://ieeexplore.ieee.org/document/10601813