10Alytics
data & strategy

Fraudulent Transactions in the Bank

10Alytics Case Study *(Data Source = Kaggle)*

Data Dictionary

- **step** - maps a unit of time in the real world. In this case 1 step is 1 hour of time. Total steps 744 (30 days simulation).
- **type** - CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.
- **amount** - amount of the transaction in local currency.
- **nameOrig** - customer who started the transaction
- **oldbalanceOrg** - initial balance before the transaction
- **newbalanceOrig** - new balance after the transaction
- **nameDest** - customer who is the recipient of the transaction
- **oldbalanceDest** - initial balance recipient before the transaction. Note that there is not information for customers that start with M (Merchants).
- **newbalanceDest** - new balance recipient after the transaction. Note that there is not information for customers that start with M (Merchants).

- **isFraud** - This is the transactions made by the fraudulent agents inside the simulation. In this specific dataset the fraudulent behavior of the agents aims to profit by taking control or customers accounts and try to empty the funds by transferring to another account and then cashing out of the system.
- **isFlaggedFraud** - The business model aims to control massive transfers from one account to another and flags illegal attempts. An illegal attempt in this dataset is an attempt to transfer more than 200.000 in a single transaction.

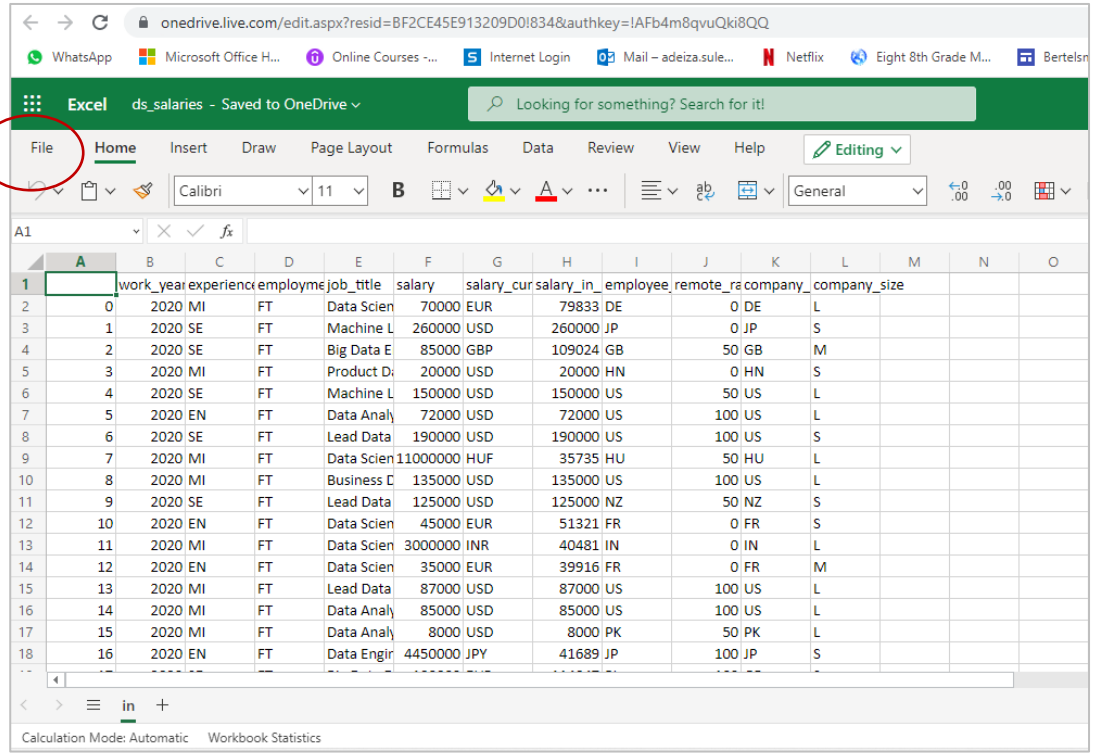
10Alytics
data & strategy

How to Download the Dataset

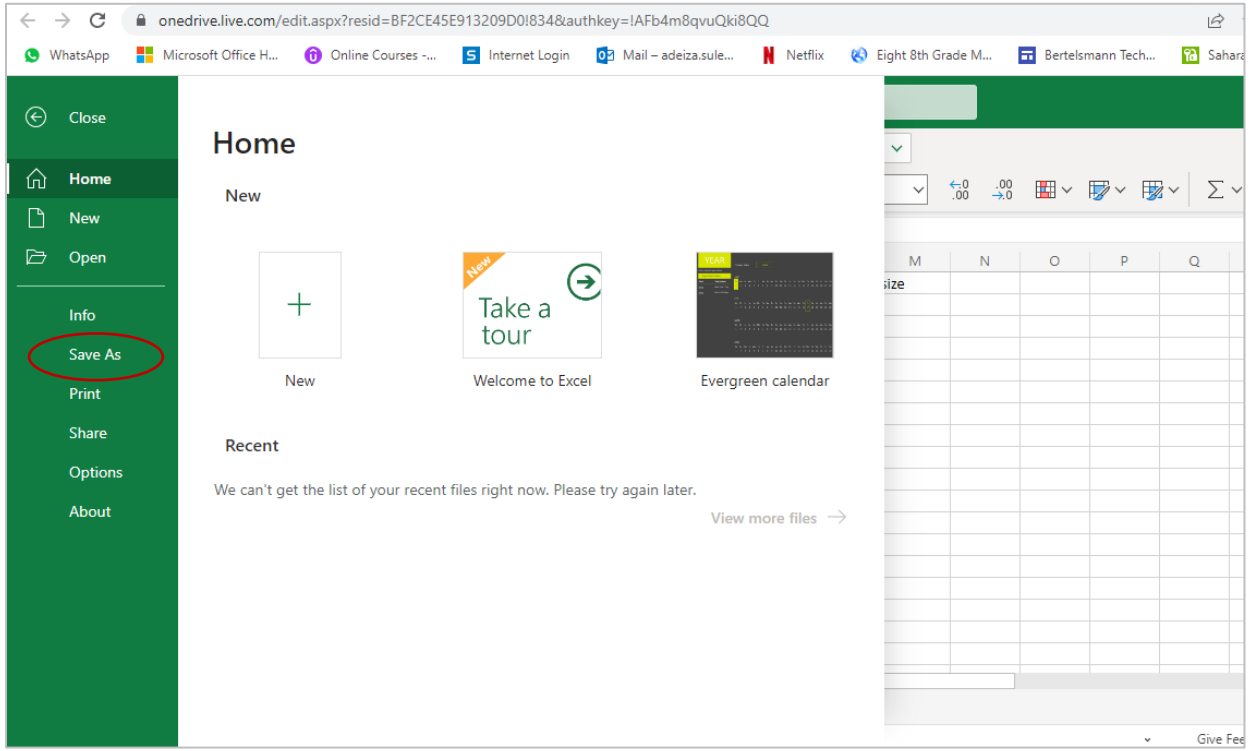
Step 1

Download your Data - HERE

Step 2



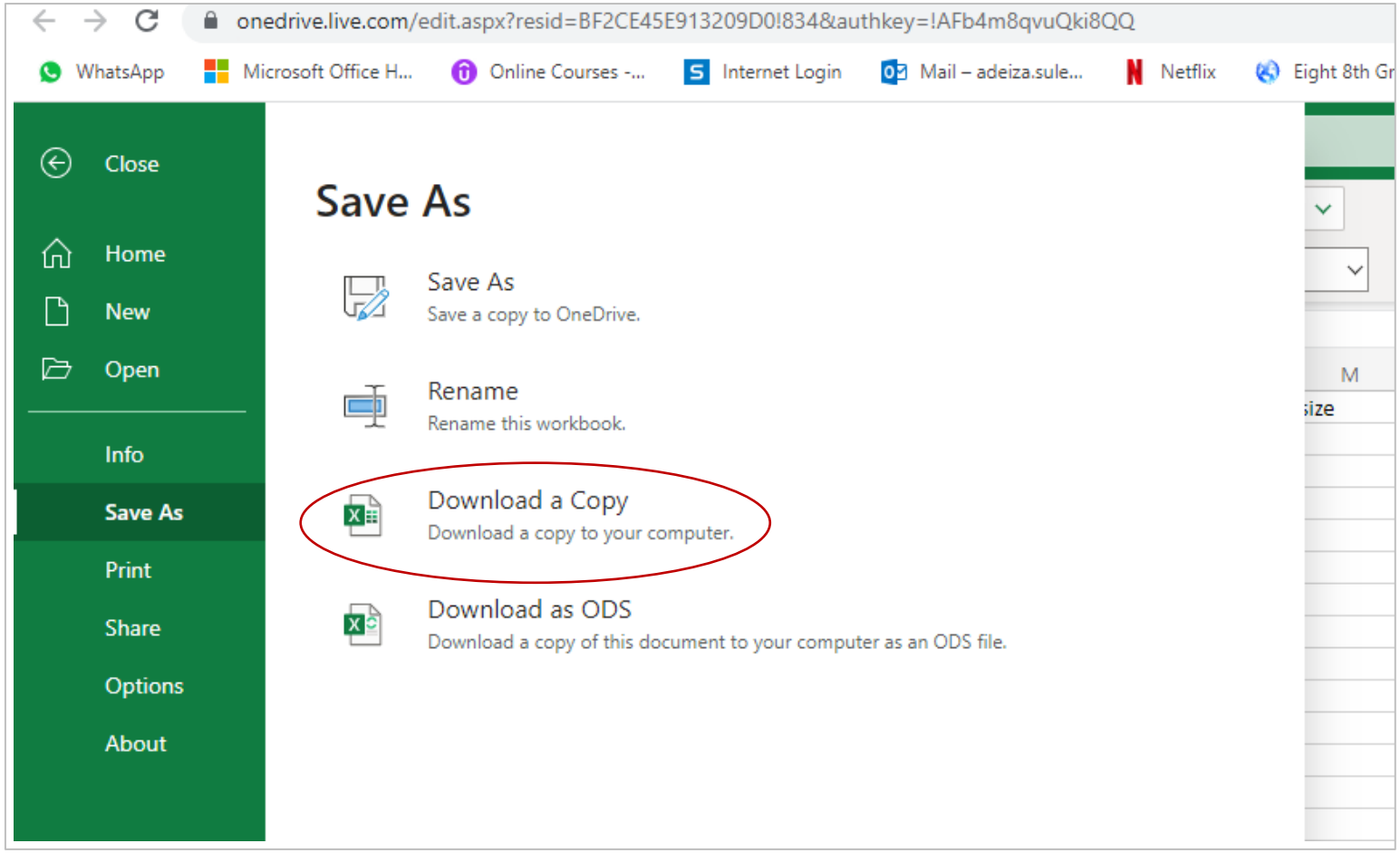
Step 3



10Alytics
data & strategy

How to Download the Dataset

Step 4



Fraudulent Transactions

The dataset in this case study contains transaction data from a bank. The data includes various fields such as step, type, amount, nameOrig, oldbalanceOrg, newbalanceOrig, nameDest, oldbalanceDest, newbalanceDest, isFraud, and isFlaggedFraud.

The step field maps a unit of time in the real world, where 1 step is equivalent to 1 hour of time, and the dataset includes a total of 744 steps (30 days simulation). The type field contains information on the type of transaction, including CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER.

Other important fields in the dataset include the amount of the transaction in local currency and information on the customer who initiated the transaction (nameOrig), as well as the customer who received the transaction (nameDest). Additionally, the dataset includes information on the initial balance of the customers' accounts (oldbalanceOrg and oldbalanceDest) and the new balance after the transaction (newbalanceOrig and newbalanceDest).

One particularly important aspect of this dataset is the presence of fraudulent transactions. The field isFraud indicates whether the transaction was made by fraudulent agents. These agents aim to profit by taking control of customer accounts and transferring funds to another account before cashing out of the system. The isFlaggedFraud field is used to identify illegal attempts to transfer more than 200.000 in a single transaction.

In this case study, your task is to explore this dataset, analyzing the patterns and characteristics of legitimate and fraudulent transactions

Tailored Analysis

Load the Data into your PostgreSQL or any other DBMS and solve the questions below:

1. How many transactions occurred per transaction type?
2. Which Transaction Type has the highest number of Fraudulent Transactions?
3. What is the average fraudulent transaction amount?
4. What is the Maximum fraudulent transaction amount?
5. What is the Minimum fraudulent transaction amount?
6. Who are the Top 10 customers with the highest amount defrauded?
7. How effective is the bank in flagging fraud?
8. Who are the Top 20 Fraudsters