



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.

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, newbalanceOrg, 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

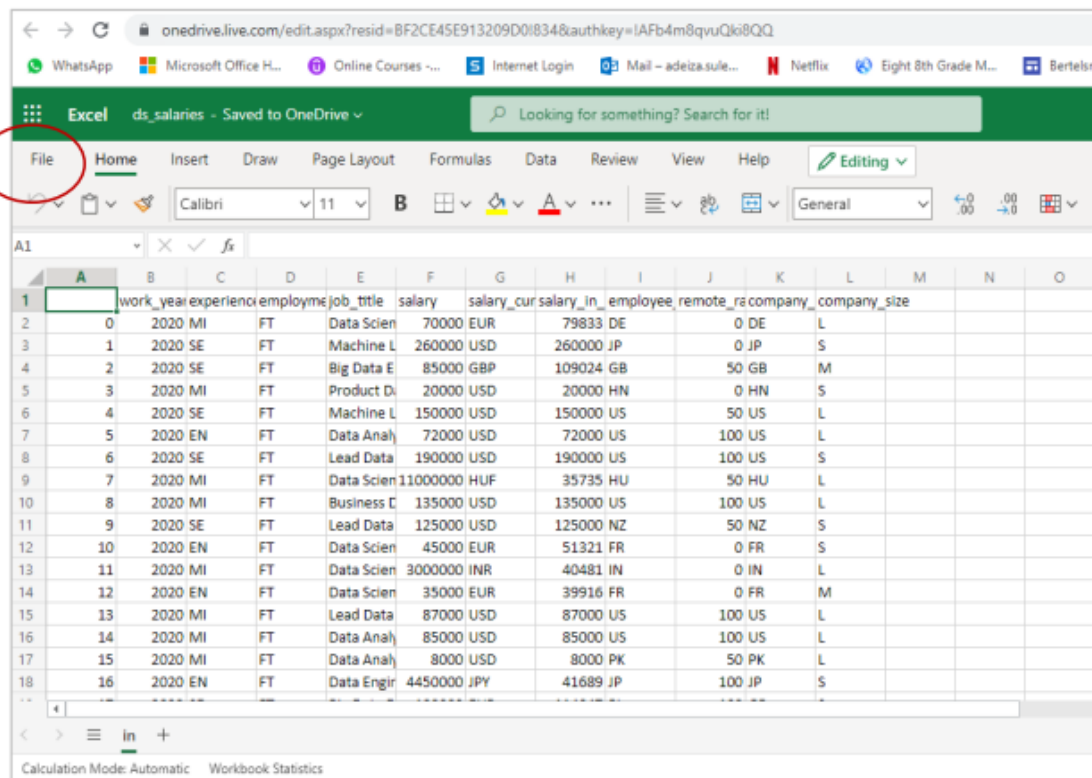


How to Download the Dataset

Step 1

Download your Data - HERE

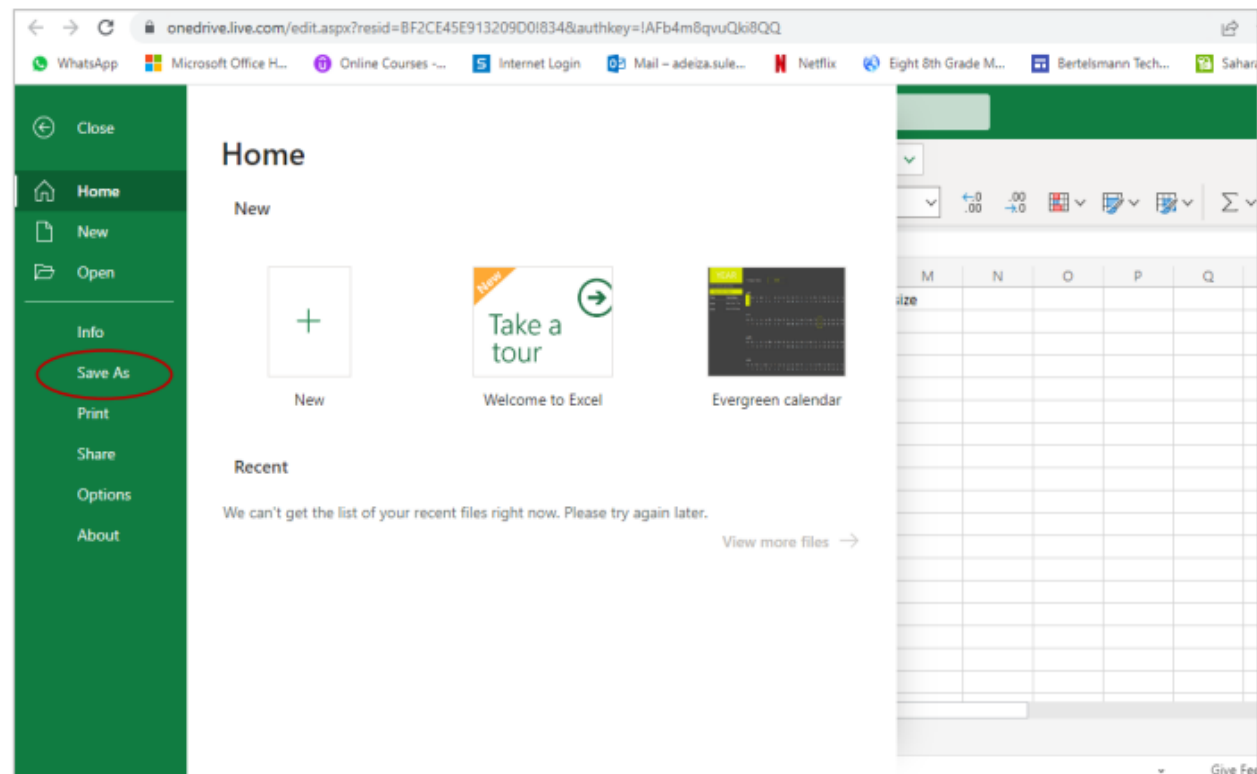
Step 2



The screenshot shows the Microsoft Excel application window. The 'File' menu is highlighted with a red circle. The ribbon at the top includes File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, and Help. The 'File' menu is open, showing options like New, Open, Save, Save As, Print, Share, Options, and About. The spreadsheet data is visible in the background.

	work_year	experience	employee	job_title	salary	salary_cur	salary_in	employee_remote_ra	company	company_size	
2	0	2020	MI	FT	Data Scien	70000	EUR	79833	DE	0 DE	L
3	1	2020	SE	FT	Machine L	260000	USD	260000	JP	0 JP	S
4	2	2020	SE	FT	Big Data E	85000	GBP	109024	GB	50 GB	M
5	3	2020	MI	FT	Product D	20000	USD	20000	HN	0 HN	S
6	4	2020	SE	FT	Machine L	150000	USD	150000	US	50 US	L
7	5	2020	EN	FT	Data Anal	72000	USD	72000	US	100 US	L
8	6	2020	SE	FT	Lead Data	190000	USD	190000	US	100 US	S
9	7	2020	MI	FT	Data Scien	1100000	HUF	35735	HU	50 HU	L
10	8	2020	MI	FT	Business C	135000	USD	135000	US	100 US	L
11	9	2020	SE	FT	Lead Data	125000	USD	125000	NZ	50 NZ	S
12	10	2020	EN	FT	Data Scien	45000	EUR	51321	FR	0 FR	S
13	11	2020	MI	FT	Data Scien	3000000	INR	40481	IN	0 IN	L
14	12	2020	EN	FT	Data Scien	35000	EUR	39916	FR	0 FR	M
15	13	2020	MI	FT	Lead Data	87000	USD	87000	US	100 US	L
16	14	2020	MI	FT	Data Anal	85000	USD	85000	US	100 US	L
17	15	2020	MI	FT	Data Anal	8000	USD	8000	PK	50 PK	L
18	16	2020	EN	FT	Data Engir	4450000	JPY	41689	JP	100 JP	S

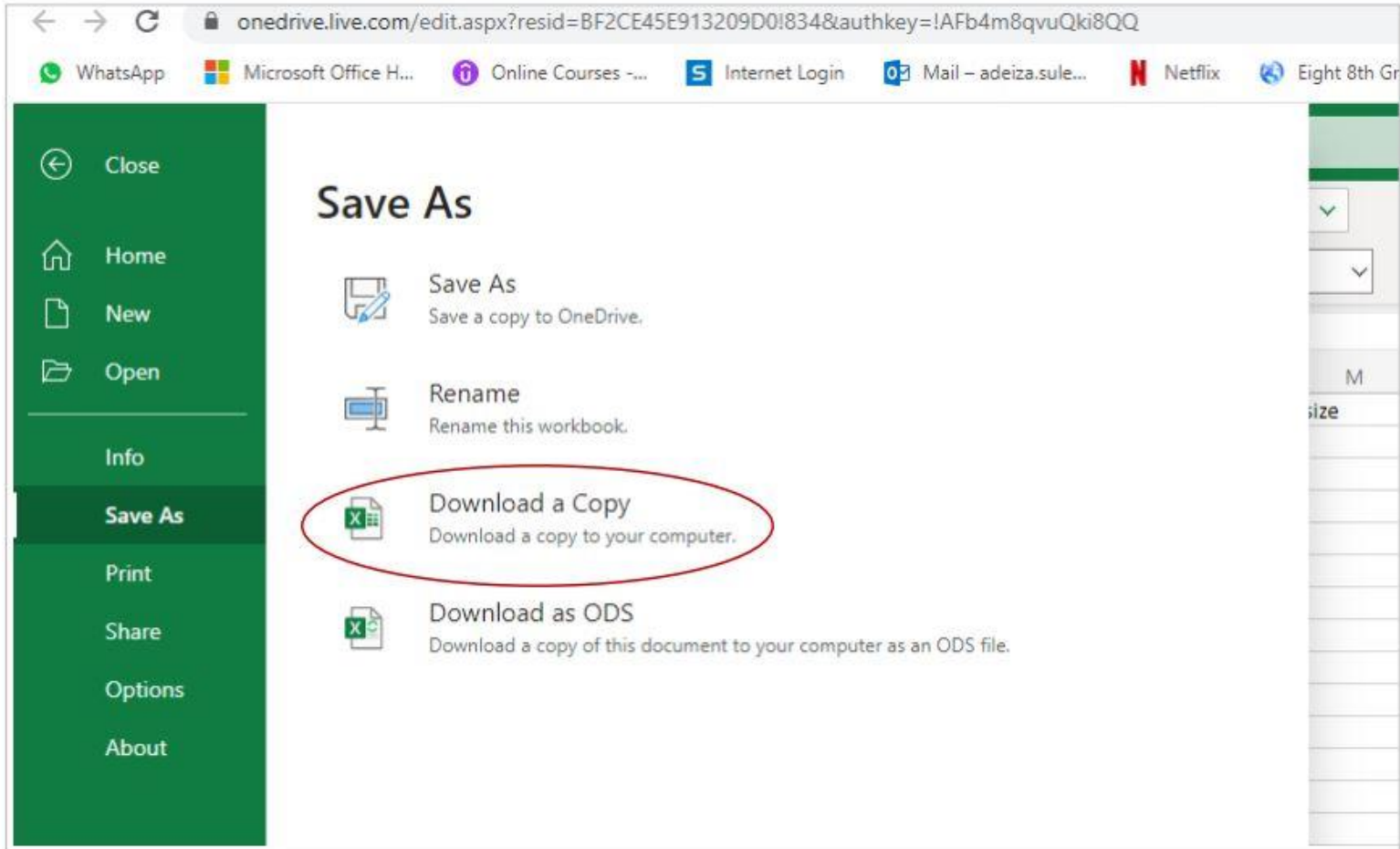
Step 3



The screenshot shows the Microsoft Excel application window with the 'File' menu open. The 'Save As' option is highlighted with a red circle. The 'Home' tab is selected in the ribbon. The 'Recent' section shows a message: 'We can't get the list of your recent files right now. Please try again later.' The 'View more files' link is visible.

How to Download the Dataset

Step 4



The screenshot shows a web browser window with the URL `onedrive.live.com/edit.aspx?resid=BF2CE45E913209D0!834&authkey=!AFb4m8qvuQki8QQ`. The browser's taskbar includes icons for WhatsApp, Microsoft Office Home, Online Courses, Internet Login, Mail, Netflix, and Eight 8th Grade. On the left, a green sidebar contains navigation options: Close, Home, New, Open, Info, **Save As**, Print, Share, Options, and About. The main content area is titled 'Save As' and lists four actions:

- Save As**: Save a copy to OneDrive.
- Rename**: Rename this workbook.
- Download a Copy**: Download a copy to your computer. (This option is circled in red in the original image.)
- Download as ODS**: Download a copy of this document to your computer as an ODS file.

INSTRUCTIONS;

- Open the excel file downloaded above and save as a '.csv' file.
- Create a Database in PostgreSQL.
- Create a Table with the descriptions below (From the guidelines provided in previous class).

```
CREATE TABLE fraudulent (  
  step INT,  
  type VARCHAR(20) ,  
  amount FLOAT,  
  nameorig VARCHAR(50),  
  oldbalanceorg FLOAT,  
  newbalanceorig FLOAT,  
  namedest VARCHAR(50),  
  oldbalancedest FLOAT,  
  newbalancedest FLOAT,  
  isfraud INT,  
  isflaggedfraud INT  
);
```

- Proceed to answer the questions below.
- You are required to submit a word document containing your queries.



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