

```
import pandas as pd
import numpy as np
from google.colab.patches import cv2_imshow
import matplotlib.pyplot as plt
from google.colab import files
uploaded = files.upload()
```

Choose Files

Creditcard.xlsx

- **Creditcard.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 61372 bytes, last modified: 2/4/2024 - 100% done

Saving Creditcard.xlsx to Creditcard.xlsx

```
import pandas as pd

# input excel file path
inputExcelFile = "Creditcard.xlsx"

# Reading an excel file
excelFile = pd.read_excel (inputExcelFile)

# Converting excel file into CSV file
excelFile.to_csv ("Creditcard.csv", index = None, header=True)

# Reading and Converting the output csv file into a dataframe object
dataframeObject = pd.DataFrame(pd.read_csv("Credit card.csv"))

# Displaying the dataframe object
dataframeObject
```

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDe
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M19797871
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M20442822
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264C
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997C
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M12307017
...
694	1	CASH_OUT	244361.81	C1191864687	0.0	0.00	C33524E
695	1	CASH_OUT	596617.87	C466032056	0.0	0.00	C1234776E
696	1	CASH_OUT	108600.78	C690822257	0.0	0.00	C1810132E
697	1	CASH_OUT	377520.94	C726212590	0.0	0.00	C17895502
698	1	CASH_OUT	60758.42	C1058822905	0.0	0.00	C1318822E

699 rows x 11 columns

```
df = pd.read_csv("Creditcard.csv")
df.head()
df.shape
```

(699, 11)

```
df.iloc[0:10,: ]
df.iloc[:, : ]
df.iloc[10:,:10]
```

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDe
10	1	DEBIT	9644.94	C1900366749	4465.0	0.00	C9976083
11	1	PAYMENT	3099.97	C249177573	20771.0	17671.03	M20965391
12	1	PAYMENT	2560.74	C1648232591	5070.0	2509.26	M9728652
13	1	PAYMENT	11633.76	C1716932897	10127.0	0.00	M8015691
14	1	PAYMENT	4098.78	C1026483832	503264.0	499165.22	M16353782
...
694	1	CASH_OUT	244361.81	C1191864687	0.0	0.00	C335246
695	1	CASH_OUT	596617.87	C466032056	0.0	0.00	C12347768
696	1	CASH_OUT	108600.78	C690822257	0.0	0.00	C18101326
697	1	CASH_OUT	377520.94	C726212590	0.0	0.00	C17895502
698	1	CASH_OUT	60758.42	C1058822905	0.0	0.00	C13188228

689 rows x 10 columns

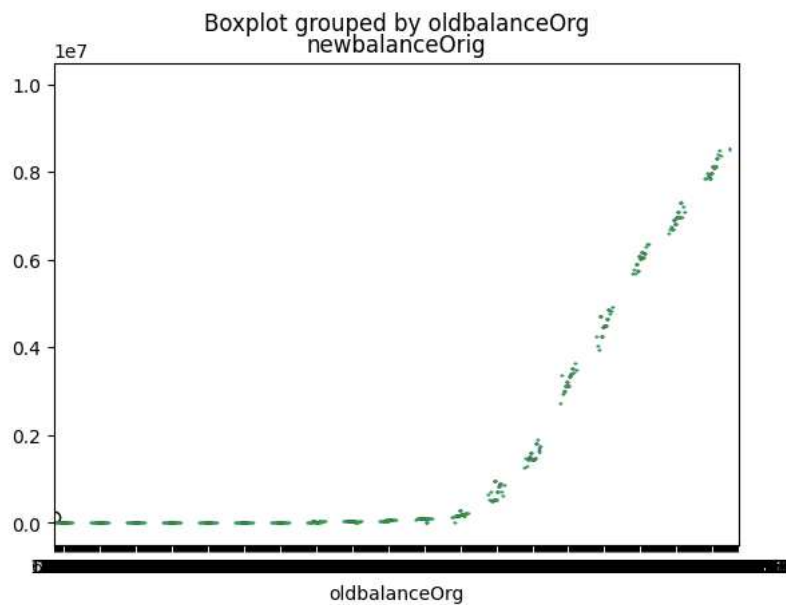
```
import numpy as np
import matplotlib.pyplot as plt
```

```
# load the dataset
df = pd.read_csv("Creditcard.csv")
```

```
# display 5 rows of dataset
df.head()
```

```
df.boxplot(by='oldbalanceOrg', column=['newbalanceOrig'], grid=False)
```

<Axes: title={'center': 'newbalanceOrig'}, xlabel='oldbalanceOrg'>



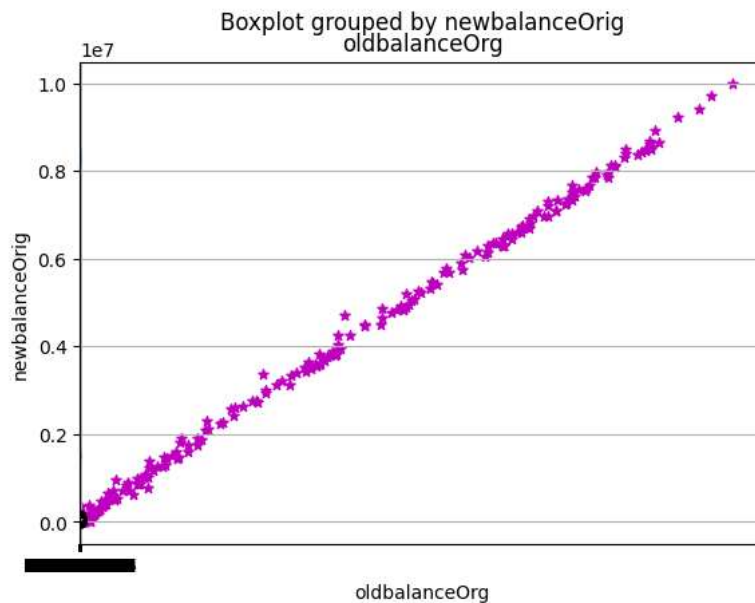
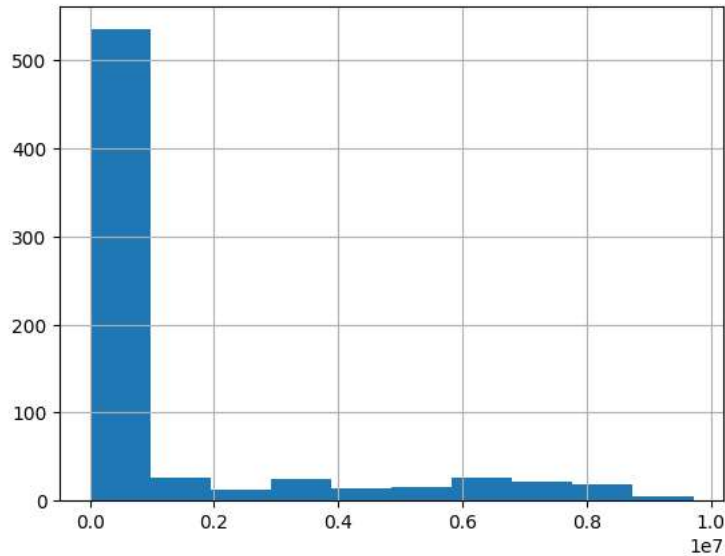
```

import matplotlib.pyplot as plt
# plot a histogram
df['oldbalanceOrig'].hist(bins=10)

# shows presence of a lot of outliers/extreme values
df.boxplot(column='oldbalanceOrig', by = 'newbalanceOrig')

# plotting points as a scatter plot
x = df["oldbalanceOrig"]
y = df["newbalanceOrig"]
plt.scatter(x, y, label= "stars", color= "m",
            marker= "*", s=30)
# x-axis label
plt.xlabel('oldbalanceOrig')
# frequency label
plt.ylabel('newbalanceOrig')
# function to show the plot
plt.show()

```



```

# making data frame from csv file
data = pd.read_csv("Creditcard.csv")

# creating bool series True for NaN values
bool_series = pd.isnull(data["newbalanceDest"])

data[bool_series]

```

