Haz doble clic (o pulsa Intro) para editar

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import pandas as pd

df = pd.read_parquet('/content/drive/MyDrive/DS4A_dataset/credit_card_data_da.parquet', engine='pyarrow')
```

```
df.info()
```

```
 7   Use Chip                          object
 8   Merchant Name                     int64
 9   Merchant City                     object
 10  Merchant State                    object
 11  Zip                               float64
 12  MCC                               int64
 13  Errors?                           object
 14  Is Fraud?                         object
 15  hour                              object
 16  minute                            object
 17  date                              datetime64[ns]
 18  datetime                          datetime64[ns]
 19  time_of_day                       object
 20  target                            int64
 21  charge_off                        float64
 22  merchant_city_rome                bool
 23  Person                            object
 24  Current Age                       int64
 25  Retirement Age                    int64
 26  Birth Year                        int64
 27  Birth Month                       int64
 28  Gender                            object
 29  Address                           object
 30  Apartment                         float64
 31  City                              object
 32  State                             object
 33  Zipcode                           int64
 34  Latitude                          float64
 35  Longitude                         float64
 36  Per Capita Income - Zipcode       float64
 37  Yearly Income - Person            float64
 38  Total Debt                        float64
 39  FICO Score                        int64
 40  Num Credit Cards                  int64
 41  personal_to_zipcode_income_diff   float64
 42  total_debt_personal_income_ratio  float64
 43  total_debt_cards_ratio            float64
 44  CARD INDEX                        int64
 45  Card Brand                        object
 46  Card Type                         object
 47  Card Number                       int64
 48  Expires                           object
 49  CVV                               int64
 50  Has Chip                          object
 51  Cards Issued                      int64
 52  Credit Limit                      float64
 53  Acct Open Date                    object
 54  Year PIN last Changed             int64
 55  Card on Dark Web                  object
 56  level_2                           int64
 57  rolling_charge_off                float64
 58  rolling_fraud_count               float64
 59  rolling_tran_count                float64
 60  rolling_tran_volume               float64
 61  transaction_count                 float64
 62  years_since_pin_change            int64
dtypes: bool(1), datetime64[ns](2), float64(18), int64(22), object(20)
memory usage: 3.2+ GB
```

```
df['Is Fraud?'].value_counts()
```

```
No     6869425
Yes       8412
Name: Is Fraud?, dtype: int64
```

```
df.isnull().sum()
```

```
User                        0
Card                        0
Year                        0
Month                       0
Day                         0
                           ...
rolling_fraud_count      8469
rolling_tran_count       8469
rolling_tran_volume      8469
transaction_count        8469
years_since_pin_change      0
Length: 63, dtype: int64
```

```
# Extract the hour and minute to perform a more refined time series analysis
df["Hour"] = df["Time"].str [0:2]
df["Minute"] = df["Time"].str [3:5]
df = df.drop(['Time'],axis=1)
```

```
# change the is fraud column to binary
df["Is Fraud?"] = df["Is Fraud?"].apply(lambda x: 1 if x == 'Yes' else 0)
```

```
fraud_data = df[df['Is Fraud?'] == 1]
```

```
# Export DataFrame to CSV and save it in Google Drive
fraud_data.to_csv('/content/drive/My Drive/Fraud_yes.csv', index=False)
```