Pyspark Case Study 1

ONLINE BANKING ANALYSIS

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, regexp_replace, trim, to_date, sum as spark_sum, max as spark_max, min as spark_min

spark = SparkSession.builder.appName("OnlineBankingAnalysis").getOrCreate()

loan_df = spark.read.option("header", "true").csv("/content/loan.csv", inferSchema=True)
credit_df = spark.read.option("header", "true").csv("/content/credit_card.csv", inferSchema=True)

txn_df = spark.read.option("header", "true").csv("/content/txn.csv", inferSchema=True)

loan_df = loan_df.selectExpr([f" '(c)' as '(c.strip().replace(' ', '_'))" for c in loan_df.columns])

credit_df = credit_df.selectExpr([f" '(c)' as '(c.strip().replace(' ', '_'))" for c in credit_df.columns])

txn_df = txn_df.selectExpr([f" '(c)' as '(c.strip().replace(' ', '_'))" for c in txn_df.columns])

loan_df = loan_df.withColumn("Loan_Amount_CLEAN", regexp_replace(trim(col("Loan_Amount")), ",", "").cast("double"))

txn_df = txn_df.withColumn("MITHDRAWAL_AMT_CLEAN", regexp_replace(trim(col("WITHDRAWAL_AMT")), ",", "").cast("double"))

txn_df = txn_df.withColumn("MITHDRAWAL_AMT_CLEAN", regexp_replace(trim(col("BALANCE_AMT")), ",", "").cast("double"))

txn_df = txn_df.withColumn("Date", to_date("Value_Date", "dd-MM-yyyyy"))
```

1. Number of loans in each category

```
# ------ LOAN DATA -----
     print("1.Number of loans in each category")
     loan_df.groupBy("Loan_Category").count().show()
1. Number of loans in each category
           Loan_Category|count|
             TRAVELLING |
BOOK STORES |
                             53 |
7 |
       AGRICULTURE |
GOLD LOAN |
EDUCATIONAL LOAN |
                             12
                              20
              AUTOMOBILE
                              60
                BUSINESS
                              24
     COMPUTER SOFTWARES
                 DTNNTNG
                              14
                SHOPPING
             RESTAURANTS
                              41
             ELECTRONICS
                             14
                BUILDING
              RESTAURANT
         HOME APPLITANCES!
                             14
```

2. Number of people who have taken more than 1 lack loan

```
print("2.People who took loan > 1 lakh")
filtered_loan 1 = loan_df.filter(col("Loan_Amount_CLEAN") > 100000)
print("Count:", filtered_loan_1.count())
filtered_loan_1.show(truncate=False )
2.People who took loan > 1 lakh
Count: 450
                    |Customer_ID|Age|Gender|Occupation
                                                                                                                                                                                                      |Marital_Status|Family_Size|Income|Expenditure|Use_Frequency|Loan_Category
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |Loan_Amount|Overdue|Debt_Record|Returned_Cheque|Dishonour_of_Bill|Loan_Amount_CLEAN|
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           10,00,000 | 5
6,00,000 | 7
2,00,000 | 2
12,09,867 | 8
3,99,435 | 9
9,87,611 | 7
                                                                   | 30 | MALE | BANK MANAGER | 29 | MALE | TEACHER | 34 | MALE | POLICE | 39 | FEMALE | TEACHER
                                                                                                                                                                                                      ISTNGLE
                                                                                                                                                                                                                                                                                                                       |50000 |22199
|45767 |12787
                                                                                                                                                                                                                                                                                                                                                                                                                                                               HOUSTNG
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |6
|0
|1
|6
|4
|8
|9
                      IB14018
                                                                                                                                                                                                         MARRIED
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 GOLD LOAN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           600000.0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |43,898
|29,999
|51,987
                      TR14022
                                                                                                                                                                                                       STNGLE
                                                                                                                                                                                                                                                                                                                        |43521 |11999
|46619 |18675
                      IB14025
                                                                                                                                                                                                          MARRIED
                                                                                                                                                                                                                                                                                                                                                                                                                                                               |HOUSING
|AUTOMOBILE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           1209867.0
                                                                                                                                                                                                                                                                                                                       45008 17454
                      IB14029
                                                                    24 FEMALE TEACHER
                                                                                                                                                                                                       SINGLE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           399435.0
                      IB14039
                                                                      45 MALE ACCOUNT MANAGER
                                                                                                                                                                                                         MARRIED
                                                                                                                                                                                                                                                                                                                        45777 | 18452
                                                                                                                                                                                                                                                                                                                                                                                                                                                                  GOLD LOAN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           39,999
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          987611.0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            5,99,934
12,90,929
1,67,654
                      IB14041
                                                                      | 59 | FEMALE | ASSISTANT PROFESSOR | MARRIED
                                                                                                                                                                                                                                                                                                                       |50999 |22999
|60111 |27111
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 EDUCATIONAL LOAD
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |599934.0
|1290929.0
                                                                    25 | FEMALE | DOCTOR
                     IB14042
                                                                25 FEMALE DOCTOR
31 | MALE | STONE KEEPER
36 | MALE | CLVIL ENGINEER
38 | FEMALE DOCTOR
38 | FEMALE DOCTOR
36 | FEMALE | CLERK
36 | FEMALE | CLERK
40 | MALE | PUBLIC WORKS
66 | FEMALE | TEACHER
38 | MALE | ELECTRICIAN
51 | FEMALE | TEACHICIAN
52 | MALE | PROFESSOR
47 | MALE | SYSTEM ENGINEER
21 | FEMALE | MANAGER
                                                                                                                                                                                                       SINGLE
                                                                                                                                                                                                                                                                                                                                                                                                                                                               TRAVELLING
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           18,000
                                                                                                                                                                                                                                                                                                                      | Ge111 | 27111 | 40999 | 11999 | 11999 | 11999 | 10999 | 60000 | 25000 | 40000 | 15000 | 35000 | 15000 | 20000 | 30000 | 15000 | 30000 | 15000 | 30000 | 15000 | 15000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 40000 | 4000
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |4,500
|19,999
|21,000
                     IB14045
                                                                                                                                                                                                       SINGLE
                                                                                                                                                                                                                                                                                                                                                                                                                                                               BOOK STORES
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          167654.0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 167654.0
| 1965577.0
| 990900.0
| 490900.0
| 490900.0
| 490900.0
| 257789.0
| 125463.0
| 1245789.0
| 654725.0
| 2569874.0
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              10,65,577
                      IB14050
IB14054
                                                                                                                                                                                                         MARRIED
MARRIED
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 HOUSING
                                                                                                                                                                                                   |MARRIED
|SINGLE
|MARRIED
|MARRIED
|MARRIED
|MARRIED
|MARRIED
|SINGLE
|SINGLE
|SINGLE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           9,00,000 | 5

4,00,000 | 8

3,00,000 | 9

2,57,789 | 4

3,54,789 | 5

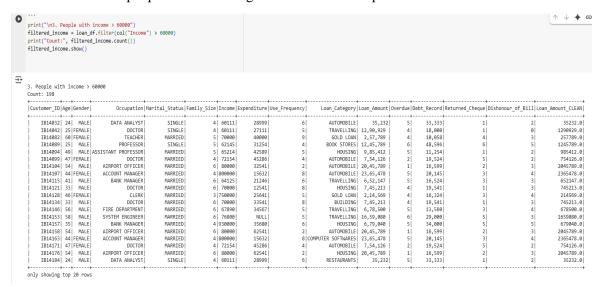
1,25,463 | 7

12,45,789 | 6

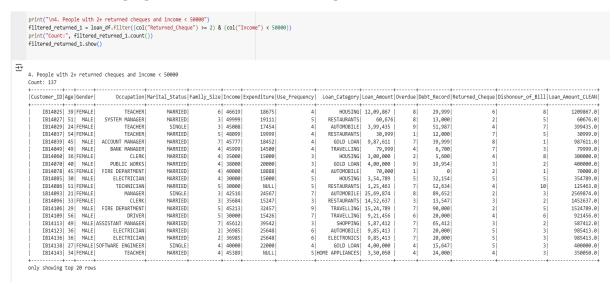
6,54,725 | 4

25,69,874 | 8
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 RESTAURANTS
                     |IB14057
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           11,111
                                                                                                                                                                                                                                                                                                                                                                                                                                                               RESTAURANTS
HOUSING
GOLD LOAN
GOLD LOAN
HOUSING
RESTAURANTS
BOOK STORES
GOLD LOAN
AUTOMOBILE
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 11,111
| 5,600
| 19,954
| 10,058
| 32,154
| 52,634
| 48,596
| 67,451
| 89,652
                    only showing top 20 rows
```

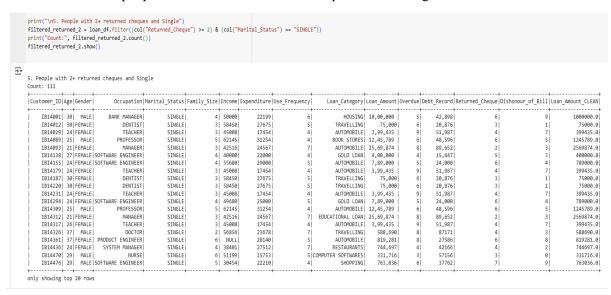
3. Number of people with income greater than 60000 rupees



4. Number of people with 2 or more returned cheques and income less than 50000



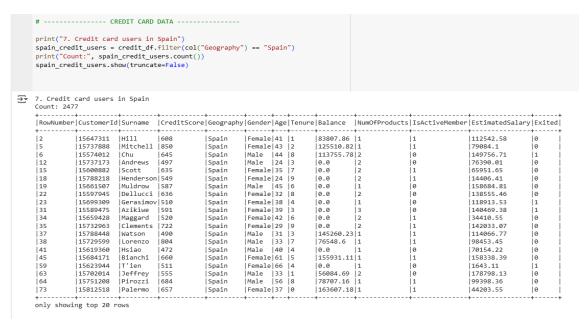
5. Number of people with 2 or more returned cheques and are single



6. Number of people with expenditure over 50000 a month

```
print("\n6. People with monthly expenses > 50000")
filtered_expense = loan_df.filter(col("Expenditure") > 50000)
 print("Count:", filtered_expense.count())
filtered_expense.show()
6. People with monthly expenses > 50000
 Customer_ID|Age|Gender| Occupation|Marital_Status|Family_Size|Income|Expenditure|Use_Frequency|Loan_Gategory|Loan_Amount|Overdue|Debt_Record|Returned_Cheque|Dishonour_of_Bill|Loan_Amount_CLEAN|
     IB14158 54 MALE AIRPORT OFFICER
IB14176 54 MALE AIRPORT OFFICER
IB14204 54 MALE AIRPORT OFFICER
IB14227 54 MALE AIRPORT OFFICER
                                                      MARRIED|
                                                                                                                    AUTOMOBILE 20,45,789
                                                      MARRIED
                                                                           6 | 80000 |
                                                                                            62541
                                                                                                                         HOUSING 20,45,789
                                                                                                                                                                16,599
                                                                                                                                                                                                                        2045789.0
                                                                                           62541
62541
                                                                                                                         DINNING 20,45,789
HOUSING 20,45,789
                                                                                                                                                               16,599
16,599
                                                      MARRTED
                                                                           6 81000
                                                                                                                                                                                                                        2045789.0
                                                      MARRIED
                                                                              80000
                                                                                                                    TRAVELLING | 6,52,147 |
HOUSING | 488,076|
      IB14278 41 MALE BANK MANAGER
                                                      MARRTED
                                                                          6 64125
                                                                                            51246
                                                                                                                                                               16.524
                                                                                                                                                                                                                         652147.0
```

7. Credit card users in Spain



8. Maximum withdrawal amount in transactions

9. Minimum withdrawal amount of an account

```
print("9.Minimum withdrawal amount")
txn_df.select(spark_min("WITHDRAWAL_AMT_CLEAN").alias("Min_Withdrawal")).show()

9.Minimum withdrawal amount
| | Min_Withdrawal|
+-----+
| 0.01|
+-----+
```

10. Maximum deposit amount of an account

11. Minimum deposit amount of an account

12. Total balance amount in every bank account

13. Customers with withdrawal amount more than 1 lakh