Questions and Results

1)Find birth country which has highest amount of people

```
>>> input4=input3.groupBy('Birth_Country').count()
>>> input5=input4.orderBy('count',ascending = False)
>>> input5.show(1)
+-------------+-----+
|Birth_Country|count|
+-------------+-----+
|        Korea|   91|
+-------------+-----+
only showing top 1 row
```

2)Find average income of people who are born in **united states of America**

```
>>> from pyspark.sql.functions import col, avg
>>> input3.filter(input3['Birth_Country'] == "United States of America").agg(avg(col("Income"))).show()
+------------------+
|       avg(Income)|
+------------------+
|208759.82352941178|
+------------------+
```

3) How many people has income over 100,000 but their loan is not approved.

```
>>> input3.filter(col("Loan_Approved") == "False").filter(col("Income") > 100000).count()
4009
```

4) Find top 10 people with highest income in **United States of america**. (Print their names, income and jobs)

```
>>> input3.select("First_Name", "Income", "Job").filter(col("Birth_Country") == "United States of America").orderBy("Income",ascending=False).show(10)
+----------+------+--------------------+
|First_Name|Income|                 Job|
+----------+------+--------------------+
|    Alyssa|482588|Amenity horticult...|
|    Hunter|468946|Psychologist, pri...|
|      Rose|426115|Adult guidance wo...|
|  Danielle|389810|Furniture conserv...|
|     Terry|380410|       Meteorologist|
|     Cindy|370322|Research scientis...|
|     Scott|368913|       Art therapist|
|   Christy|355150|      Engineer, land|
|     Kelly|341448|          Press sub|
|  Kristina|338804|          Herbalist|
+----------+------+--------------------+
only showing top 10 rows
```

5) How many number of distinct jobs are there?

```
>>> print(input3.select("Job").distinct().count())
639
```

6) How many writers earn less than 100,000?

```
>>> input3.filter(col("Job") == "Writer").filter(col("Income") < 100000).count()
5
```