

```
from pyspark.sql import functions as f
from pyspark.sql import SparkSession
df=SparkSession.builder.appName('raghav').getOrCreate()
```

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df1=df.read.csv('/FileStore/Employment_detais.csv',header=True,inferSchema=True)
df2=df.read.csv('/FileStore/loan_details.csv',header=True,inferSchema=True)
df3=df.read.csv('/FileStore/personal_details.csv',header=True,inferSchema=True)
```

```
df1.show()
df2.show()
df3.show()
```

3	Female	Yes	3+	Graduate
4	Male	Yes	1	Not Graduate
5	Male	No	0	Not Graduate
6	Male	Yes	2	Graduate
7	Male	No	0	Graduate
8	Male	Yes	0	Not Graduate
9	Male	Yes	2	Not Graduate
10	Male	Yes	0	Graduate
11	Male	Yes	0	Graduate
12	Male	No	0	Graduate
13	Female	Yes	0	Not Graduate
14	Male	No	0	Graduate
15	Male	Yes	2	Graduate
16	Male	Yes	2	Not Graduate
17	Male	Yes	1	Graduate
18	Male	Yes	1	Graduate
19	Female	No	2	Graduate
20	Male	Yes	0	Graduate

only showing top 20 rows

```
join_df = df1.join(df2, df1.Id == df2.Id, how='outer') \
.join(df3, df1.Id == df3.Id, how='outer') \
.drop(df2.Id)
```