Mini Project III:

Patterns in Banking Behavior

Ali Bahrami & Chunshan Feng

November 16, 2021

Goal

Determine which client segments are bringing in the most revenues and risks for the bank

- How do banks make money?
    - Net Interest Income: on interests they earn by lending borrowed money to clients
    - Interchange income: through fees paid by merchants per transaction
- How do banks lose money?
    - Default on loans (i.e. credit card loans, mortgage loans, business loans, etc.)

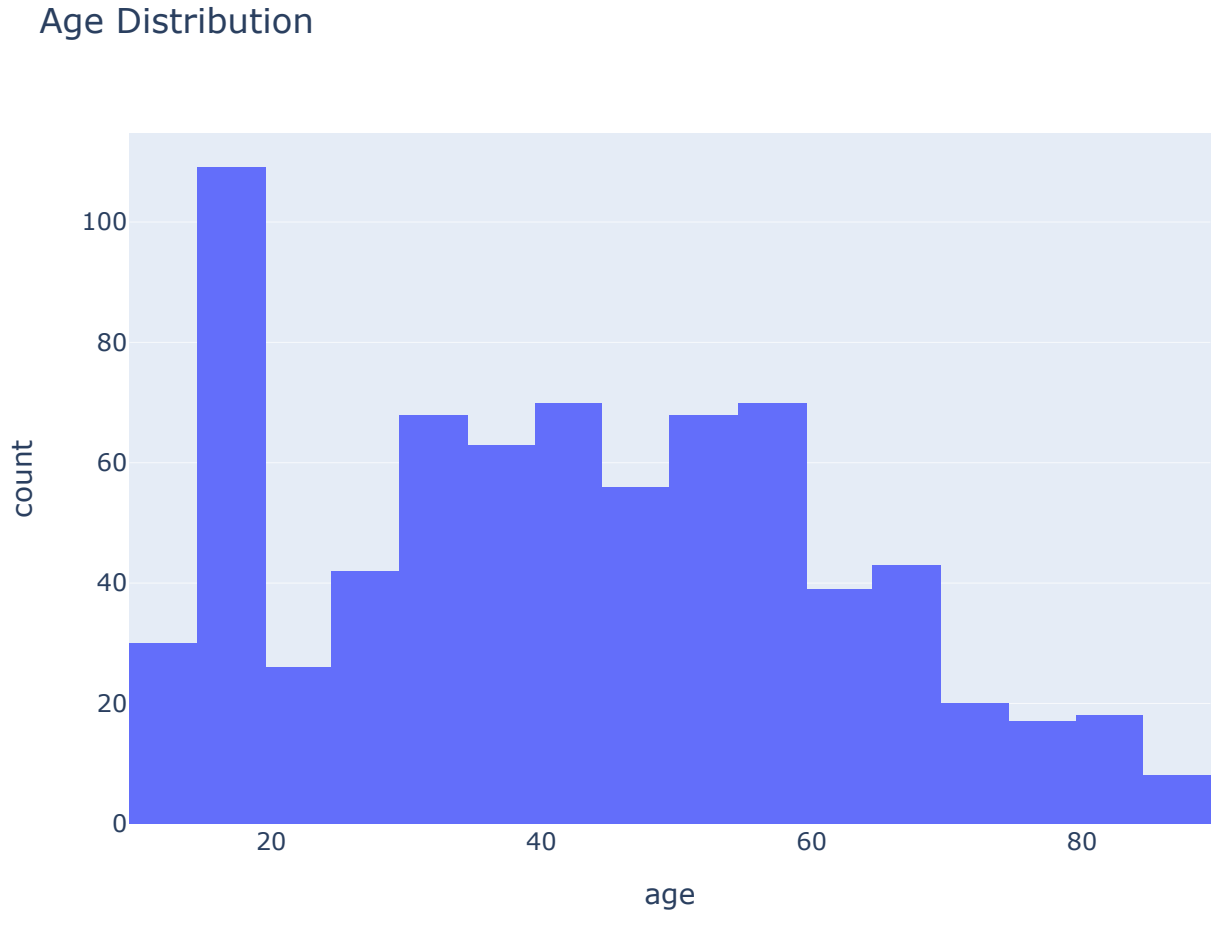
Agenda

1. Do an expolartory analysis of the bank's customer base using demographics data:
    - Income
    - Gender
    - Number of children
2. Customer segmentation using income in the following segments:
    - **revenue:** total interests paid, credit balance owed
    - **revenue:** total transactions made
    - **risk:** percent of credit spent
3. PCA Analysis and Radar Charts

Explore the Dataset

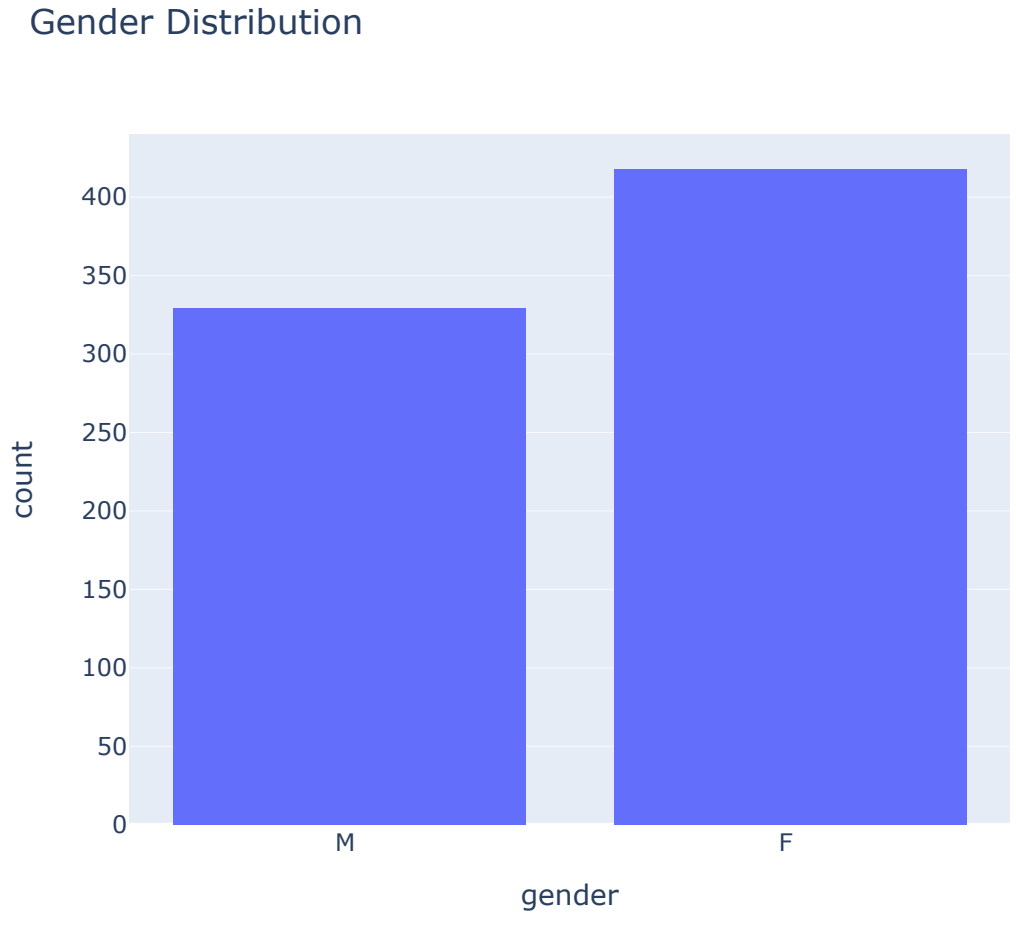
In [4]:

```
# Age Distribution
fig = px.histogram(df_customer.age, x="age", title='Age Distribution', nbins=16)
fig.show()
```



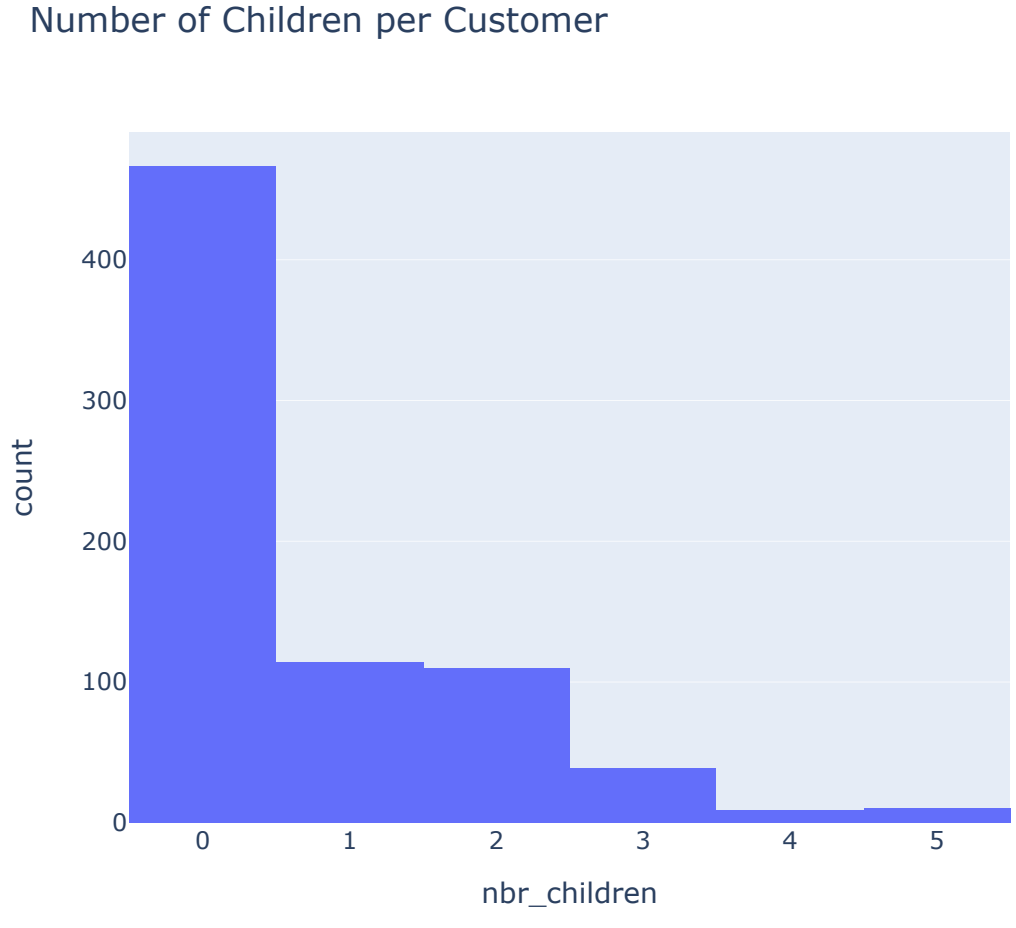
In [5]:

```
# Gender Distribution
fig = px.histogram(df_customer.gender, x="gender", title='Gender Distribution', width = 600)
fig.show()
```



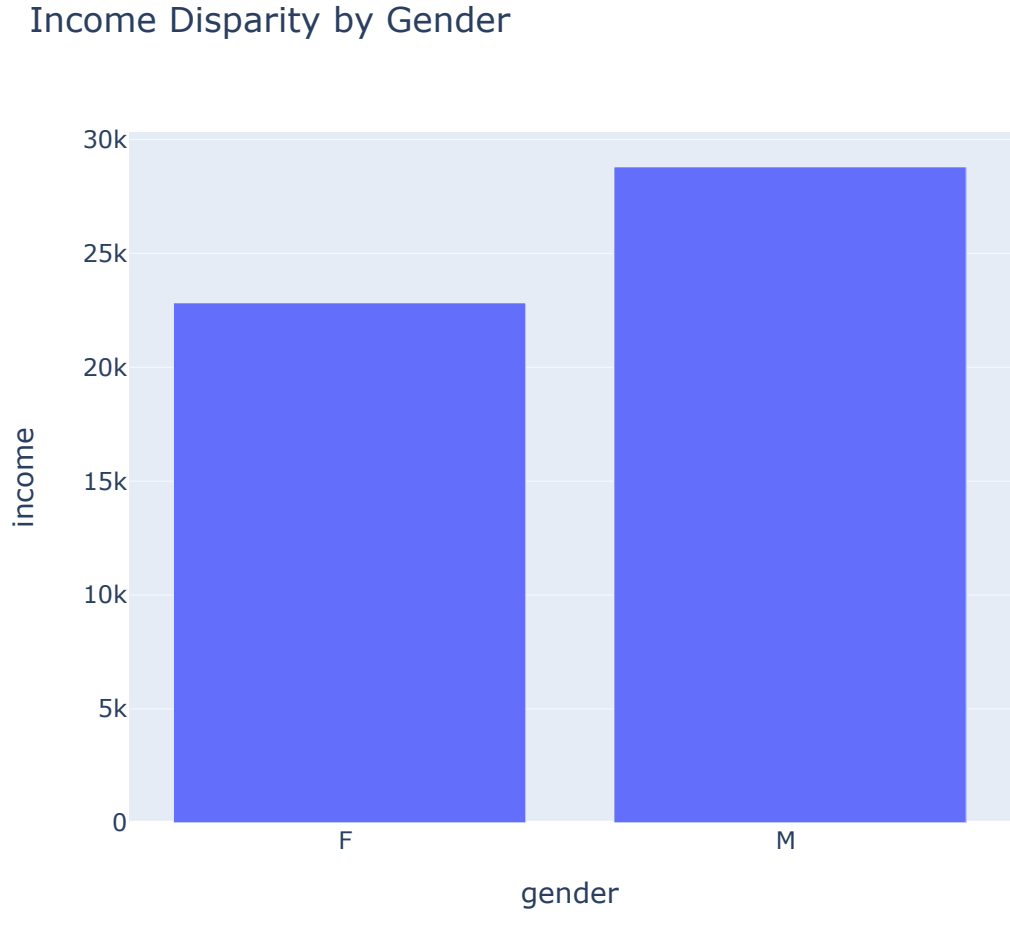
In [6]:

```
# Number of children
fig = px.histogram(df_customer.nbr_children, x="nbr_children", title='Number of Children per
fig.show()
```

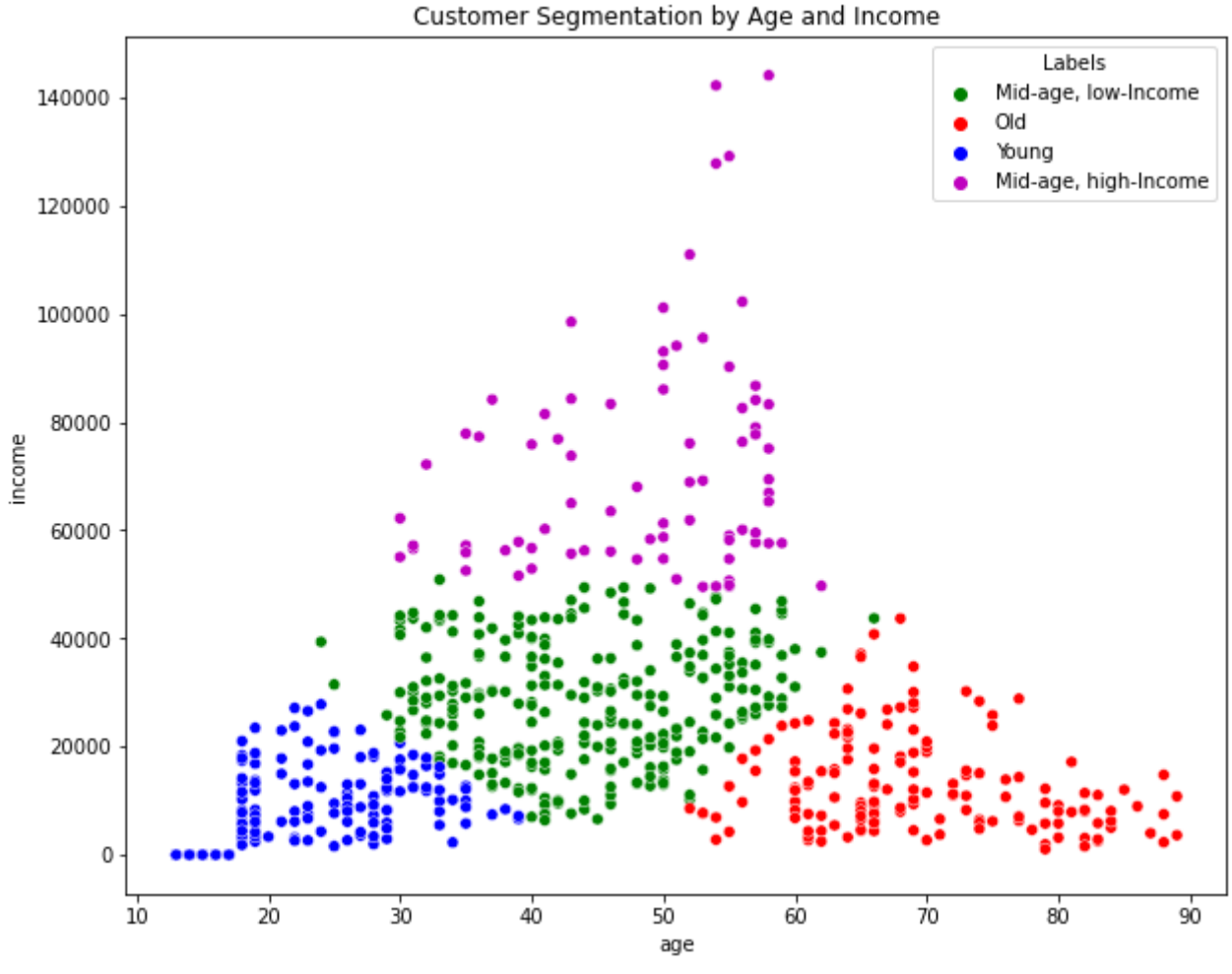


In [8]:

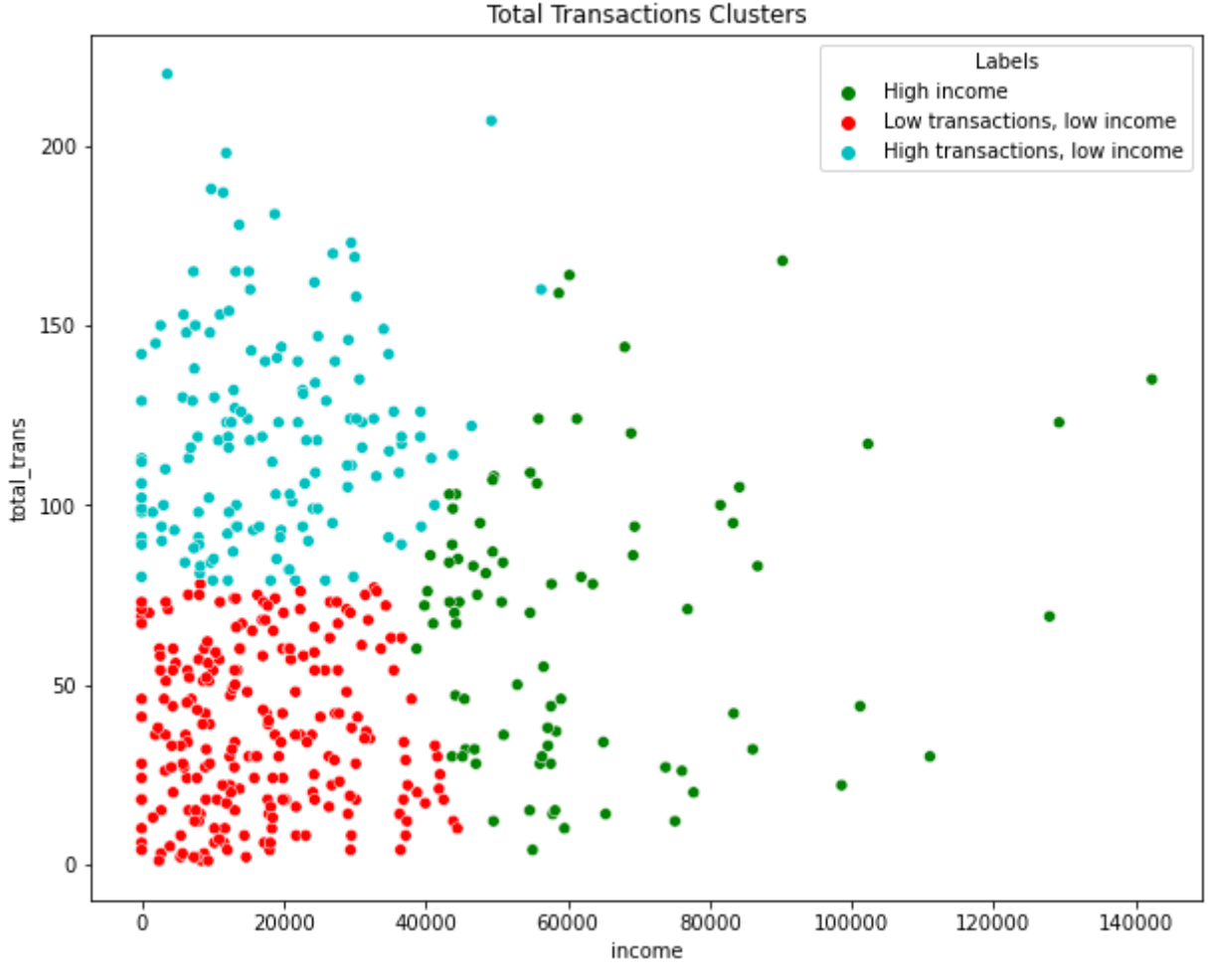
```
# Income disparity by Gender
fig.show()
```



Income Distribution by Age and Income



Clusters by Income and Transaction Amount

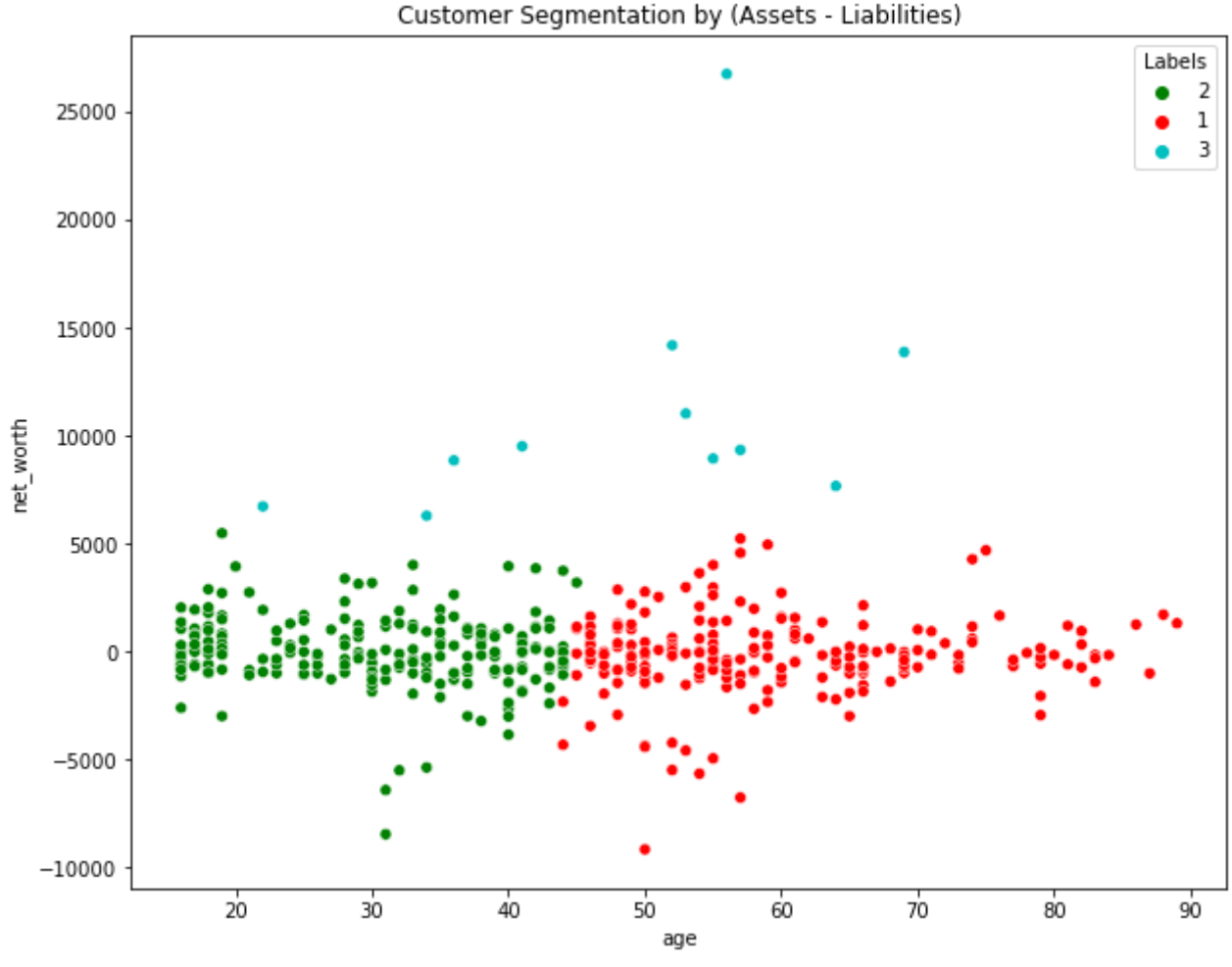


Group 3 (turquoise): Rewards products

Clusters using Credit Balance

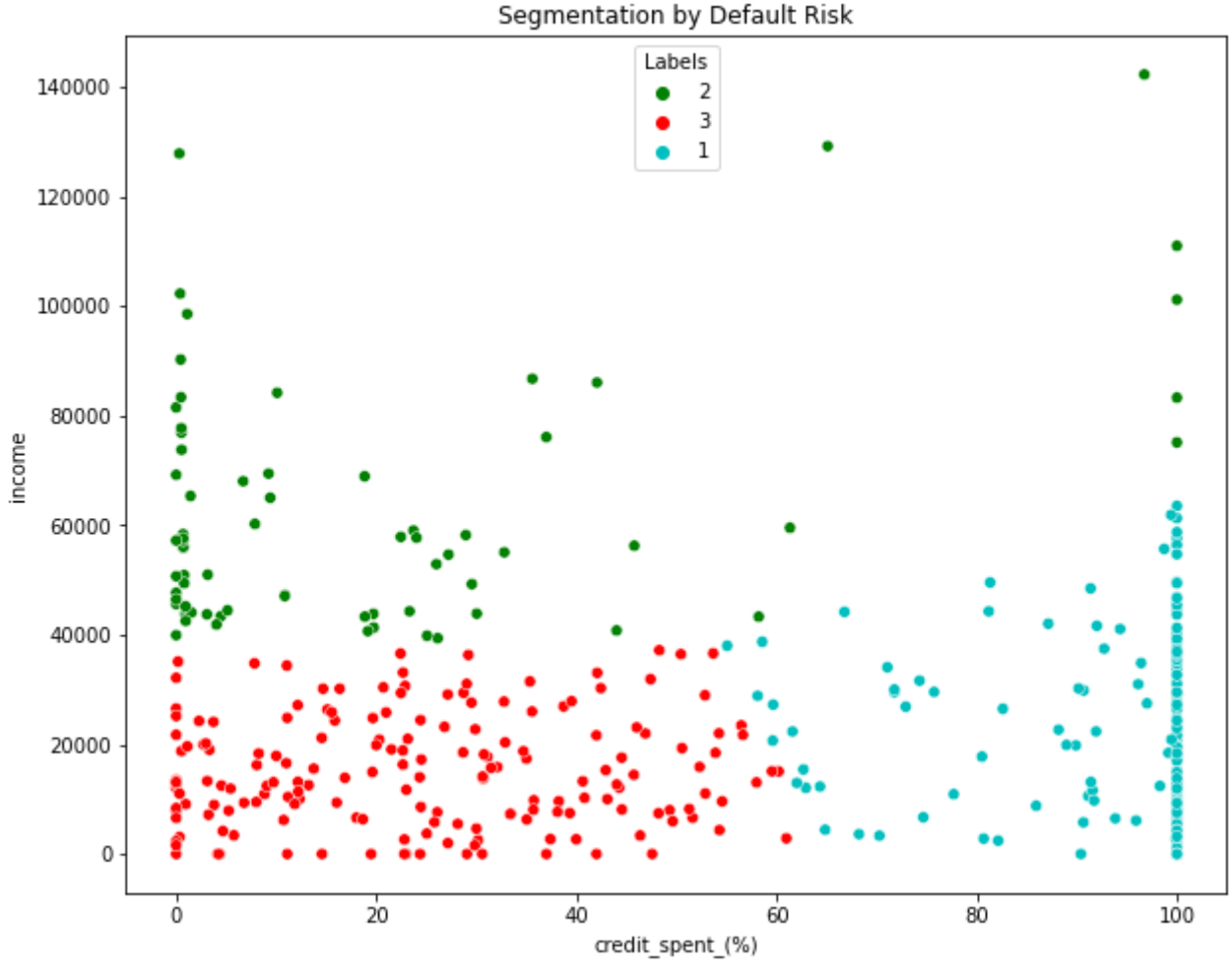
Group 3 (turquoise): Premium Cashback and Rewards Card with first-year fee rebate

Customer Segments - Net Worth



Net Worth = (Checking + Savings) - Credit

Customer Segments - Risk



Group 1 (turquoise): Credit limit increase offer

PCA Analysis

In [55]:

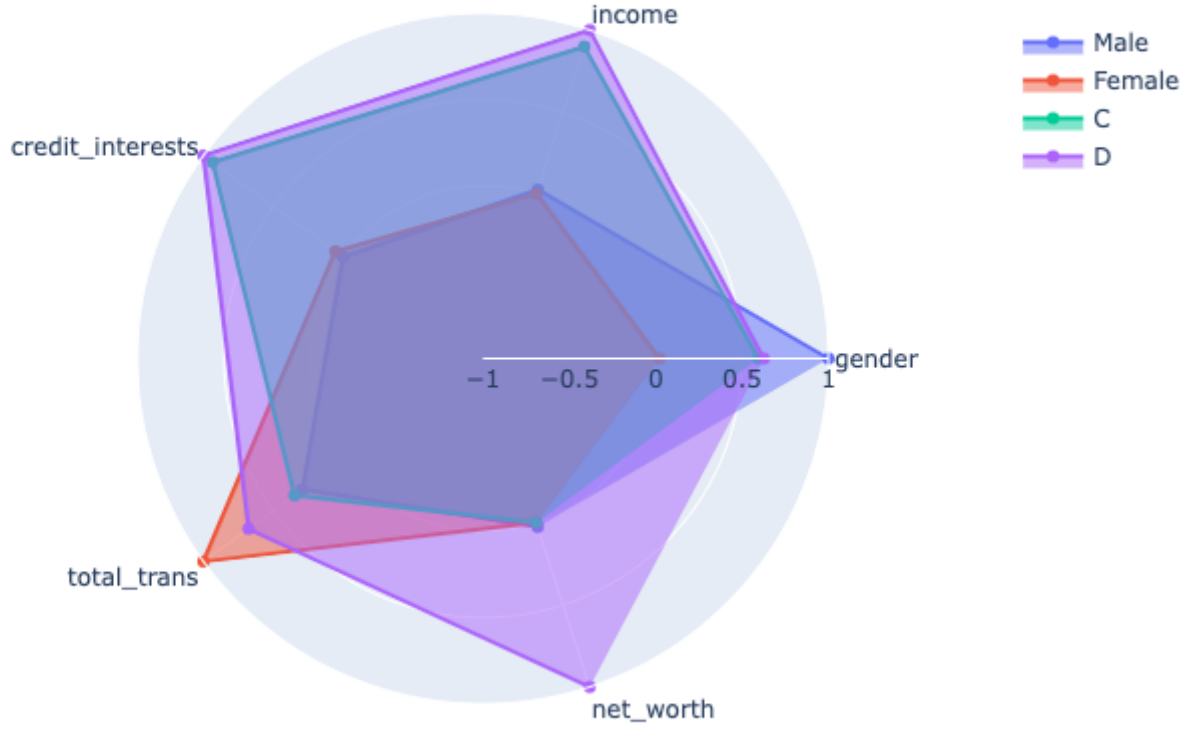
```
# Multi-dimensional Analysis
df_segmentation
```

Out[55]:

	gender	income	credit_interests	total_trans	net_worth
0	1	50890	361.26	84	1254.83
1	0	10053	14.81	54	1847.77
2	0	22690	56.93	94	-1054.51
3	1	6605	13.63	15	-134.13
4	0	55888	248.77	124	479.08
...	...	...	...	...	...
452	1	9271	79.61	62	-596.25
453	0	10244	12.21	130	-175.02
454	0	19863	41.27	18	-842.24
455	0	39942	0.00	17	0.00
456	1	142274	963.28	135	-5646.76

457 rows × 5 columns

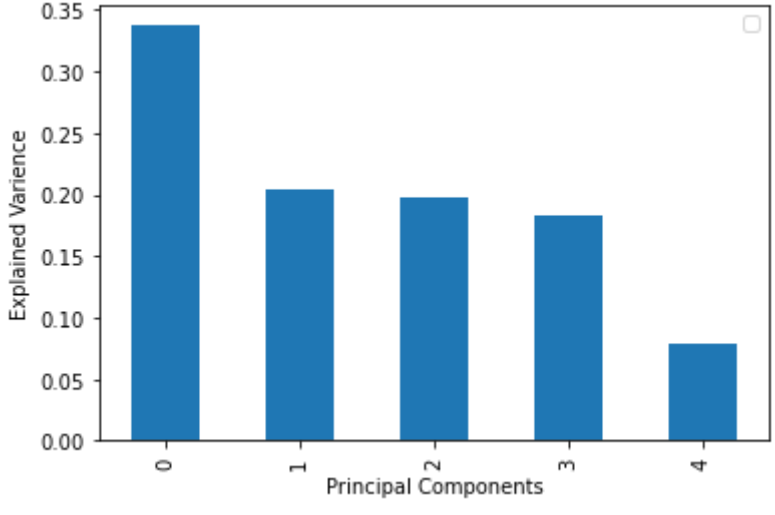
Radar Chart


income
credit_interests
gender
total_trans
net_worth
−1
−0.5
0
0.5
1
Male
Female
C
D

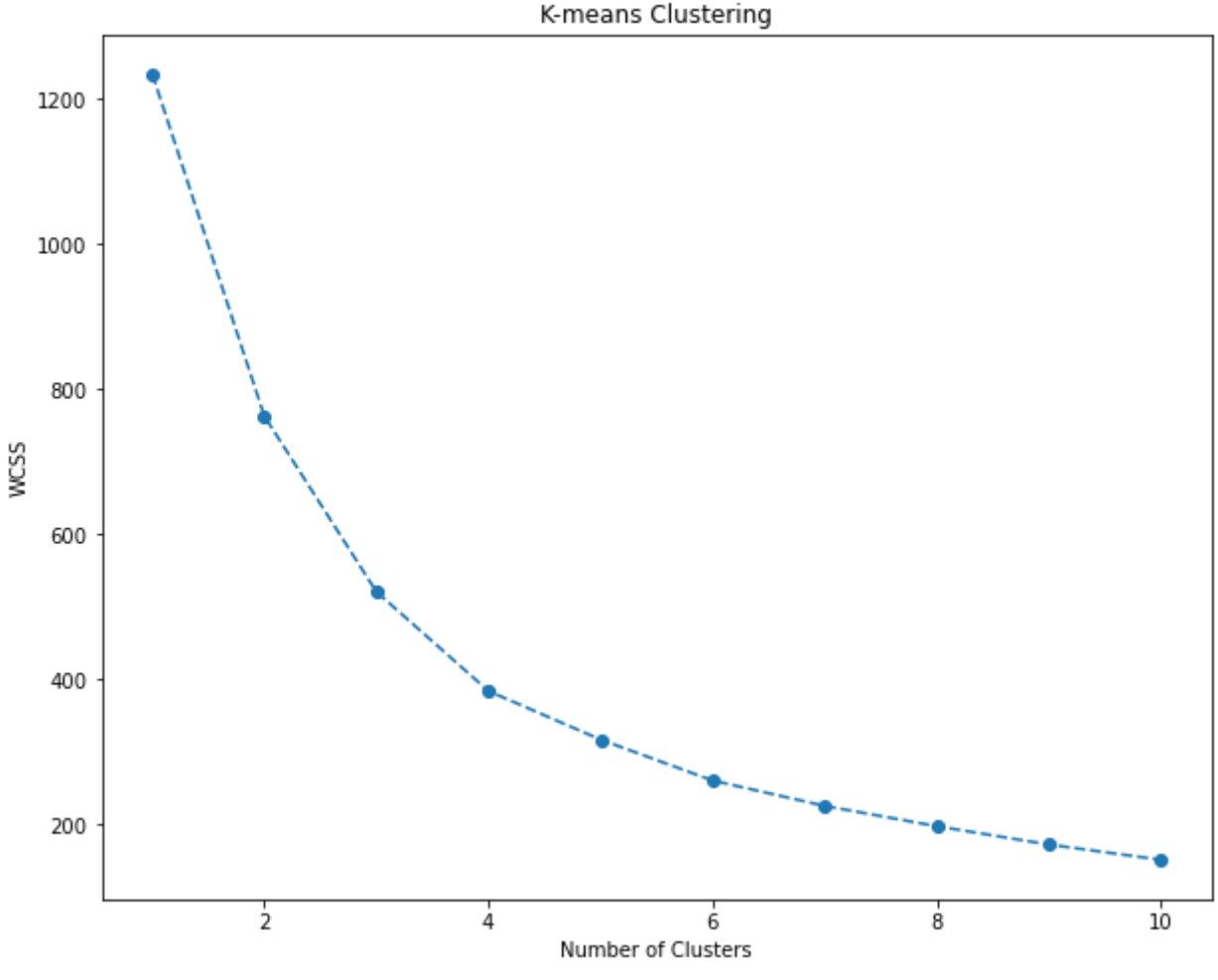
Explained Variance

Explained Variance

In [71]:
```
pd.DataFrame(pca.explained_variance_ratio_).plot.bar()
plt.legend('')
plt.xlabel('Principal Components')
plt.ylabel('Explained Varience');
```



Post-PCA K-Means Clustering



Post-PCA Cluster