Introduction

- Company XYZ which has provided us with a performance summary with some of their loan performance summarized by merchant. In the dataset there are 143 rows and 12 columns. The 12 columns are as followed:
    - merchant_id (Unique identifier for the merchant)
    - actual_repayment_pct (actual percentage of loan volume that is repaid)
    - predicted_repayment_pct (predicted percentage of loan volume that is repaid)
    - num_trxn (number of loans per merchant)
    - avg_auth_amt (total amount of user requests / apply for)
    - avg_loan_amt (total amount of the loan that we issue to the user)
    - avg_fico (score that measures a user's risk, higher score means less risk (range from 300-850))
    - avg_term (Duration of the loan)
    - avg_apr (annual percentage rate (interest rate that charged to the user))
    - name (name of the merchant)
    - category (Merchant's industry)
    - subcategory (Merchant's industry)

Company XYZ Questions

- Based on the data, what would you say drives the variance between actual and predicted repayment?
- What could drive the difference between auth_amount and loan_amount?
- Based on the analysis, in which areas would you increase or decrease volume?
- Which categories generate the most profit for Company XYZ? And which categories have the highest average loan amount?
- Would you find the top 3 merchants with the most unit profit in each category?
- What information can Company XYZ gather to further evaluate their merchants' profit with Affirm? What types of analysis, evaluation, or diligence should Company XYZ do?

Methods Used for Analysis

Python using Jupyter Notebooks

Rstudio using R Programming

Based on the data, what would you say drives the variance between actual and predicted repayment?

- Here, I compared the variances of predicted and actual repayment. The P Value here was 2.2e-16, which is less than the significance level of 0.05. This means there is a significance difference between the two variance.

```
	F test to compare two variances

data:  merged_all$predicted_repayment_pct and merged_all$actual_repayment_pct
F = 0.066226, num df = 142, denom df = 142, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.04759814 0.09214362
sample estimates:
ratio of variances 
        0.06622586 
```

Cont.

- Next, I applied Linear Regression to see which factors had a significant role when it came to predicted repayment pct. Which here, we can see that the average fico score is significant.

```
Call:
lm(formula = predicted_repayment_pct ~ avg_loan_amt + avg_auth_amt
    num_trxn + avg_fico + avg_term + avg_apr, data = merged_all)

Residuals:
      Min        1Q    Median        3Q       Max
-0.240240 -0.007117  0.008141  0.016893  0.050988

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.462e-01  8.537e-02   8.740 7.71e-15 ***
avg_loan_amt  3.076e-05  1.819e-05   1.691   0.0930 .
avg_auth_amt -2.132e-05  1.562e-05  -1.365   0.1744
num_trxn      4.146e-06  5.342e-06   0.776   0.4390
avg_fico      2.902e-04  1.218e-04   2.382   0.0186 *
avg_term      7.926e-04  1.989e-03   0.398   0.6909
avg_apr      -6.363e-02  5.342e-02  -1.191   0.2357
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0396 on 136 degrees of freedom
Multiple R-squared:  0.1832,    Adjusted R-squared:  0.1471
F-statistic: 5.082 on 6 and 136 DF,  p-value: 9.77e-05
```

Cont.

- I also applied Linear Regression to actual repayment pct. Which here, we can see that the average fico score is significant as well.

```
Call:
lm(formula = actual_repayment_pct ~ avg_loan_amt + avg_auth_amt +
    num_trxn + avg_fico + avg_term + avg_apr, data = merged_all)

Residuals:
     Min       1Q   Median       3Q      Max
-0.89568 -0.02189  0.03890  0.08180  0.19726

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.635e-02  3.516e-01  -0.103  0.91781
avg_loan_amt -7.288e-05  7.489e-05  -0.973  0.33214
avg_auth_amt  6.766e-05  6.431e-05   1.052  0.29464
num_trxn      7.685e-06  2.200e-05   0.349  0.72735
avg_fico      1.353e-03  5.016e-04   2.698  0.00787 **
avg_term      7.205e-03  8.192e-03   0.880  0.38068
avg_apr      -3.208e-02  2.200e-01  -0.146  0.88428
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.163 on 136 degrees of freedom
Multiple R-squared:  0.0827,	Adjusted R-squared:  0.04223
F-statistic: 2.044 on 6 and 136 DF,  p-value: 0.06401
```

Cont.

- Lastly, I applied ANOVA to both regressions. On the top, we can see that the response variable, being predicted repayment, we can see that average loan amount, average auth amount, and average fico are significant. While for the bottom ANOVA, the response was actual repayment and it only had average fico to be significant. In conclusion, we can see what drives the variance between predicted and actual repayment mainly is the fico score of a merchant.

```
Analysis of Variance Table

Response: predicted_repayment_pct
              Df   Sum Sq   Mean Sq F value    Pr(>F)
avg_loan_amt   1 0.024727 0.0247272 15.8491 0.0001101 ***
avg_auth_amt   1 0.008575 0.0085748  5.4961 0.0204753 *
avg_fico       1 0.010868 0.0108680  6.9659 0.0092562 **
Residuals    139 0.216862 0.0015602
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Analysis of Variance Table

Response: actual_repayment_pct
           Df Sum Sq  Mean Sq F value  Pr(>F)
avg_fico    1 0.2625 0.262493   10.06 0.00186 **
Residuals 141 3.6791 0.026093
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What could drive the difference between auth_amount and loan_amount?

- On top, we have a linear regression model for average loan amount being the dependent variable. We saw that average fico score and average term were significant factors. On the bottom, we have another linear regression model that has average auth amount being the dependent variable. Here we have average fico and average term as significant factors. Those two factors were in both models, meaning that they could drive the difference between auth amount and loan amount.

```
Call:
lm(formula = avg_loan_amt ~ actual_repayment_pct + predicted_repayment_pct +
    num_trxn + avg_fico + avg_term + avg_apr, data = merged_all)

Residuals:
    Min      1Q  Median      3Q     Max
-1601.0  -424.9   -74.1   236.5  4474.8

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -1.120e+04  1.910e+03  -5.867 3.22e-08 ***
actual_repayment_pct    -8.893e+01  4.434e+02  -0.201    0.841
predicted_repayment_pct  2.854e+03  1.804e+03   1.582    0.116
num_trxn                -7.317e-02  1.102e-01  -0.664    0.508
avg_fico                 1.223e+01  2.267e+00   5.393 2.97e-07 ***
avg_term                 1.980e+02  3.746e+01   5.287 4.84e-07 ***
avg_apr                 -9.826e+02  1.104e+03  -0.890    0.375
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 816.2 on 136 degrees of freedom
Multiple R-squared:  0.3833,    Adjusted R-squared:  0.3561
F-statistic: 14.09 on 6 and 136 DF,  p-value: 1.977e-12
```

```
Call:
lm(formula = avg_auth_amt ~ actual_repayment_pct + predicted_repayment_pct +
    num_trxn + avg_fico + avg_term + avg_apr, data = merged_all)

Residuals:
    Min      1Q  Median      3Q     Max
-1739.2  -475.5  -137.0   224.2  6102.3

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -1.076e+04  2.232e+03  -4.822 3.75e-06 ***
actual_repayment_pct     7.122e+01  5.182e+02   0.137    0.891
predicted_repayment_pct  2.429e+03  2.109e+03   1.152    0.252
num_trxn                -7.793e-02  1.288e-01  -0.605    0.546
avg_fico                 1.169e+01  2.650e+00   4.412 2.07e-05 ***
avg_term                 2.256e+02  4.378e+01   5.154 8.77e-07 ***
avg_apr                 -1.044e+03  1.290e+03  -0.810    0.420
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 953.9 on 136 degrees of freedom
Multiple R-squared:  0.3272,    Adjusted R-squared:  0.2975
F-statistic: 11.02 on 6 and 136 DF,  p-value: 5.432e-10
```

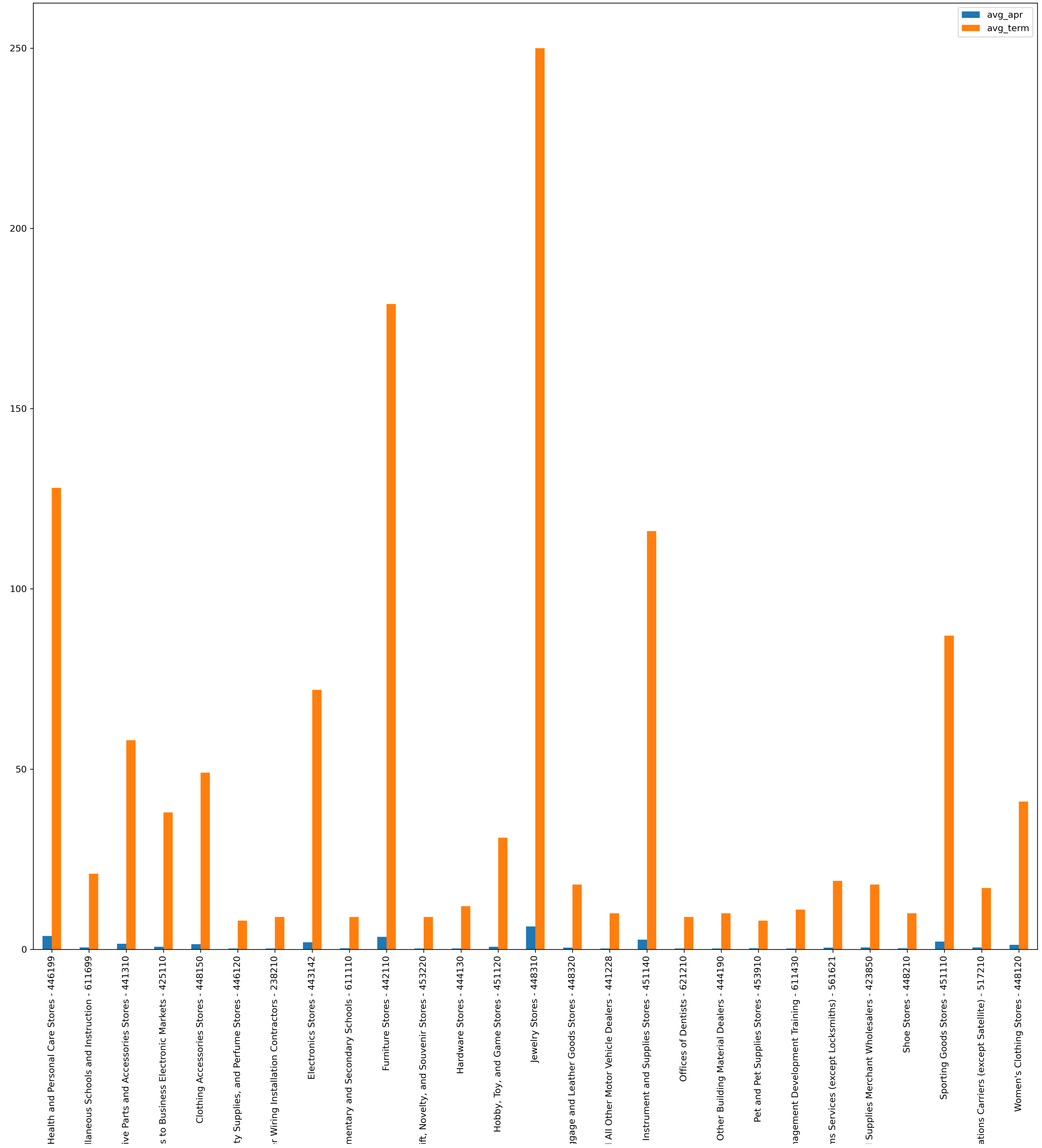
Cont.

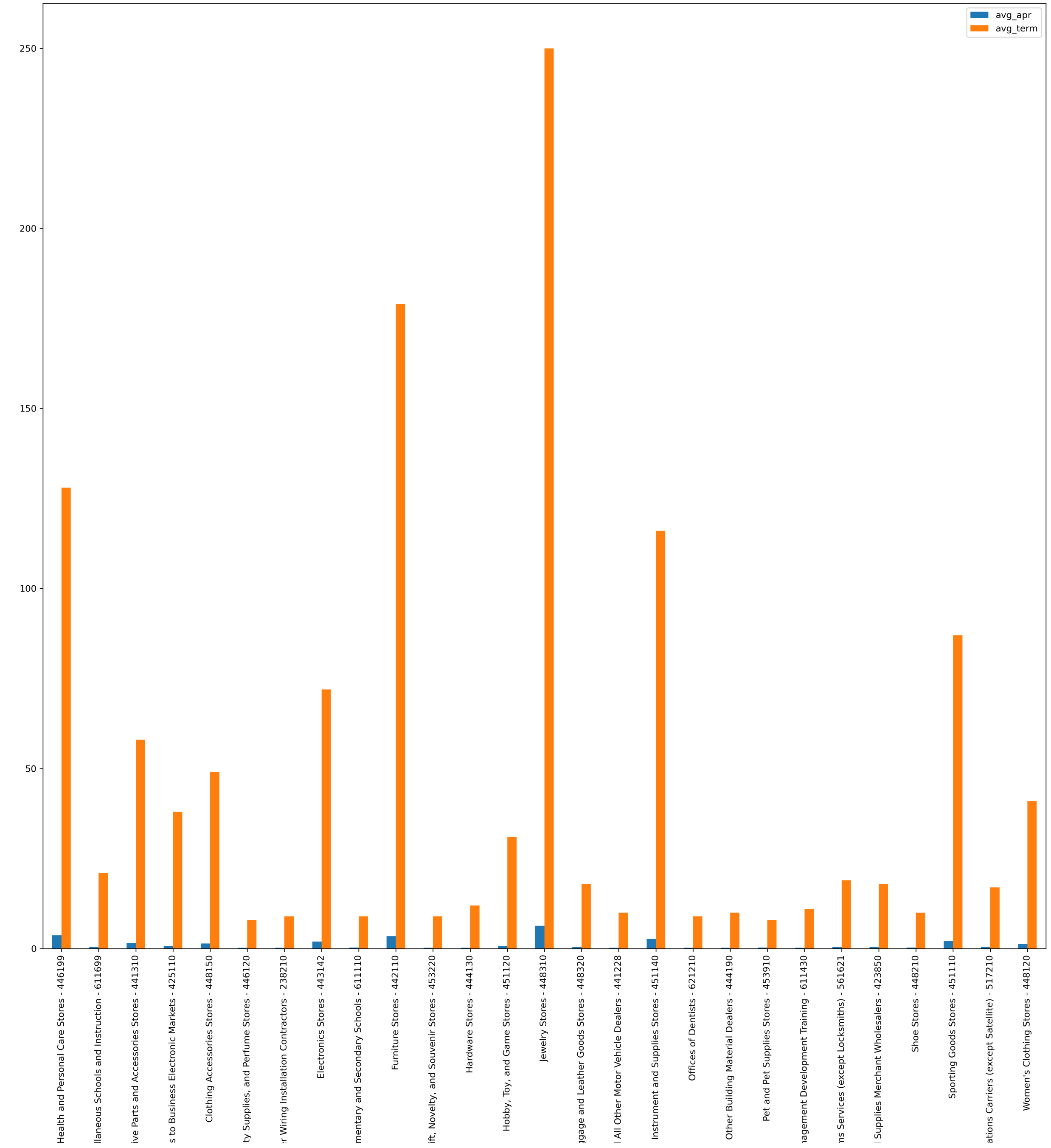
- I decided to do a correlation graph to see the correlation between the factors. Since we were looking at auth amount and loan amount, we can see that auth amount with average term has a correlation of 0.4, which is positive. But auth amount with average fico has a correlation of 0.41, which is positive as well. For loan amount, with average fico there was a positive correlation of 0.47 and loan with term it has a positive of 0.4. Both average term and average fico had a high correlation between loan and auth amount when looked at, compared to the other variables that were not loan and auth amount.

	actual_repayment_pct	predicted_repayment_pct	num_trxn	avg_auth_amt	avg_loan_amt	avg_fico	avg_term
actual_repayment_pct	1	0.34	0.02	0.17	0.17	0.26	0.1
predicted_repayment_pct	0.34	1	0.06	0.27	0.31	0.37	0.1
num_trxn	0.02	0.06	1	-0.08	-0.08	0.01	-0.14
avg_auth_amt	0.17	0.27	-0.08	1	0.98	0.41	0.4
avg_loan_amt	0.17	0.31	-0.08	0.98	1	0.47	0.4
avg_fico	0.26	0.37	0.01	0.41	0.47	1	0.04
avg_term	0.1	0.1	-0.14	0.4	0.4	0.04	1

Which categories generate the most profit for Company XYZ?

- A company usually generates profit through APR and loan term length. The longer a term is, the client usually will end up paying more in APR. As well if the APR is higher too. Blue is APR and orange is term. Here we have the subcategory of Jewelry having the longest-term length and the highest APR out of the other subcategories. The second most profitable is Furniture Stores.

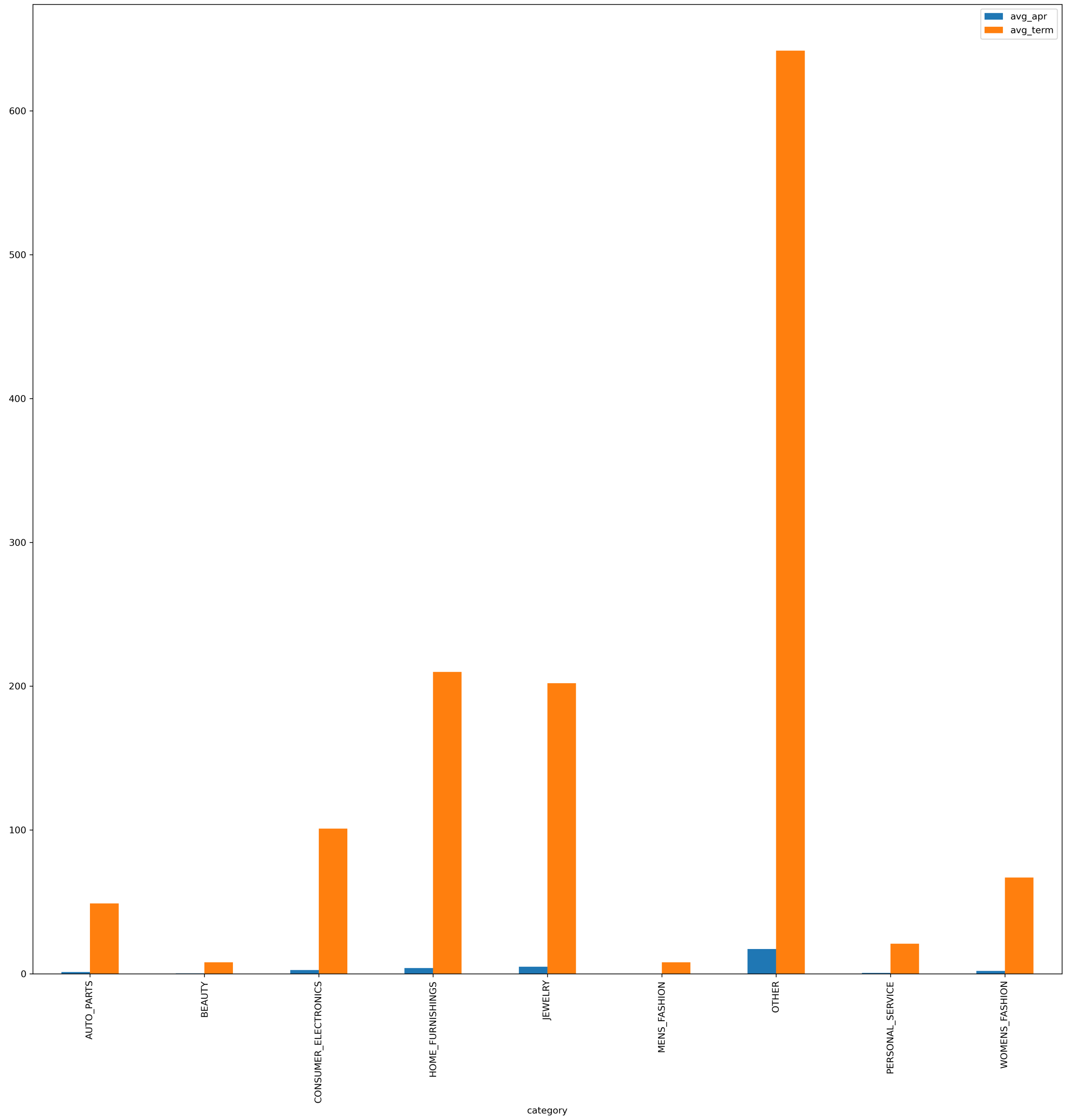




avg_apr
avg_term
0
50
100
150
200
250
Health and Personal Care Stores - 446199
llaneous Schools and Instruction - 611699
ive Parts and Accessories Stores - 441310
s to Business Electronic Markets - 425110
Clothing Accessories Stores - 448150
ty Supplies, and Perfume Stores - 446120
r Wiring Installation Contractors - 238210
Electronics Stores - 443142
mentary and Secondary Schools - 611110
Furniture Stores - 442110
ift, Novelty, and Souvenir Stores - 453220
Hardware Stores - 444130
Hobby, Toy, and Game Stores - 451120
Jewelry Stores - 448310
ggage and Leather Goods Stores - 448320
All Other Motor Vehicle Dealers - 441228
Instrument and Supplies Stores - 451140
Offices of Dentists - 621210
Other Building Material Dealers - 444190
Pet and Pet Supplies Stores - 453910
agement Development Training - 611430
ns Services (except Locksmiths) - 561621
Supplies Merchant Wholesalers - 423850
Shoe Stores - 448210
Sporting Goods Stores - 451110
ations Carriers (except Satellite) - 517210
Women's Clothing Stores - 448120

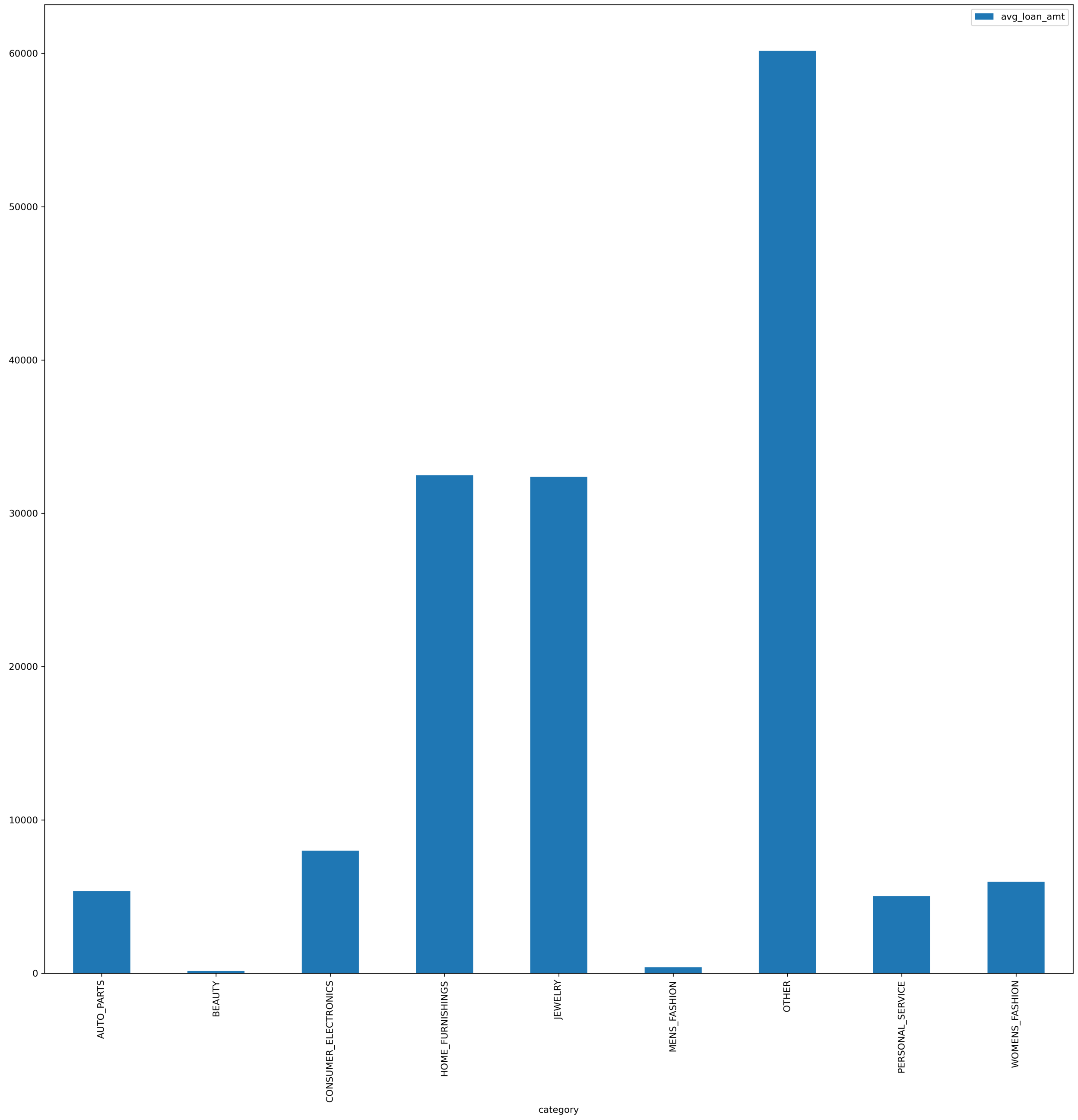
Cont.

- Blue is APR and orange is term. Here we have the category of Other having the longest-term length and the highest APR out of the other categories. The second most profitable is Home Furnishings.



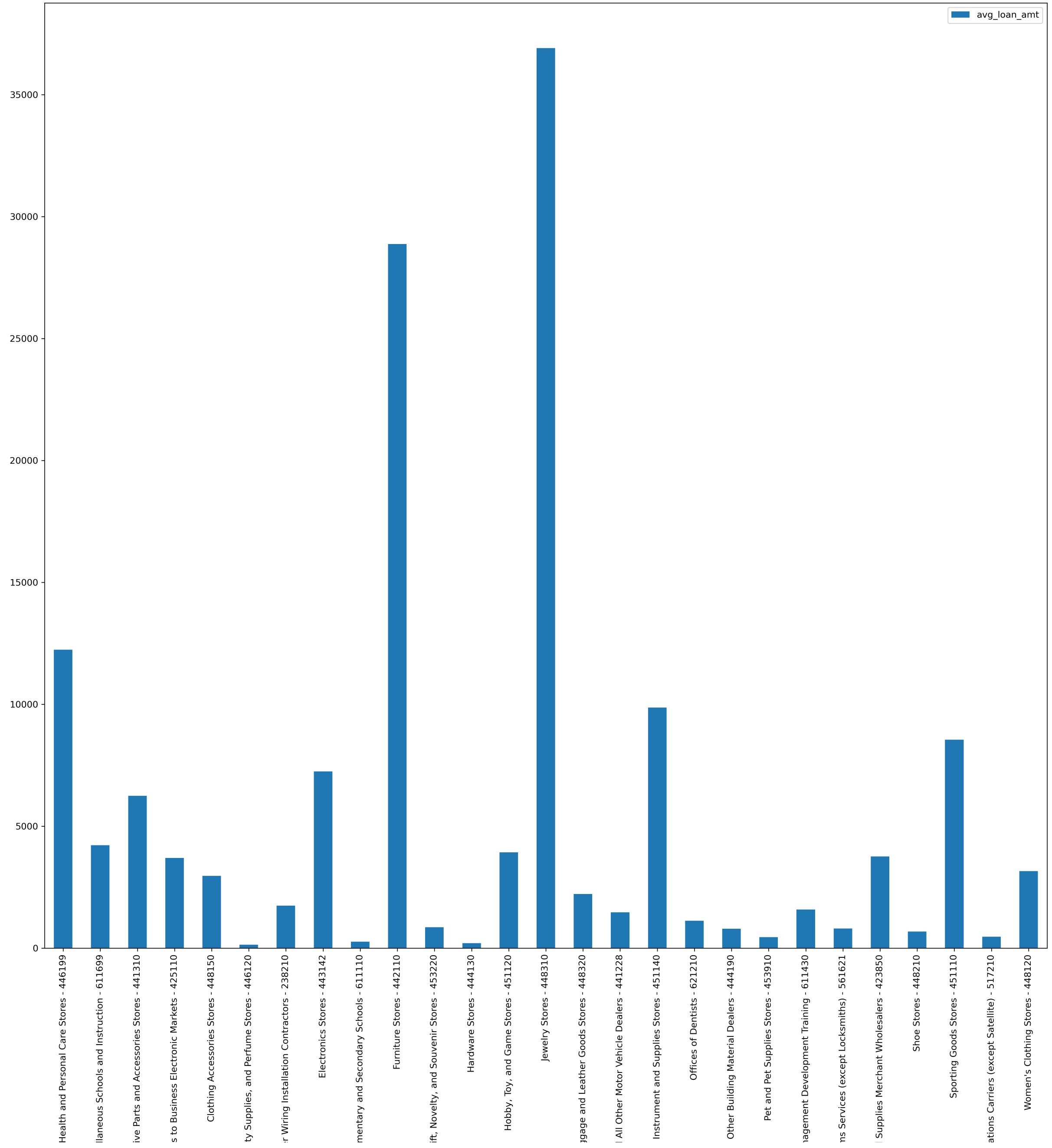
Which categories have the highest average loan amount?

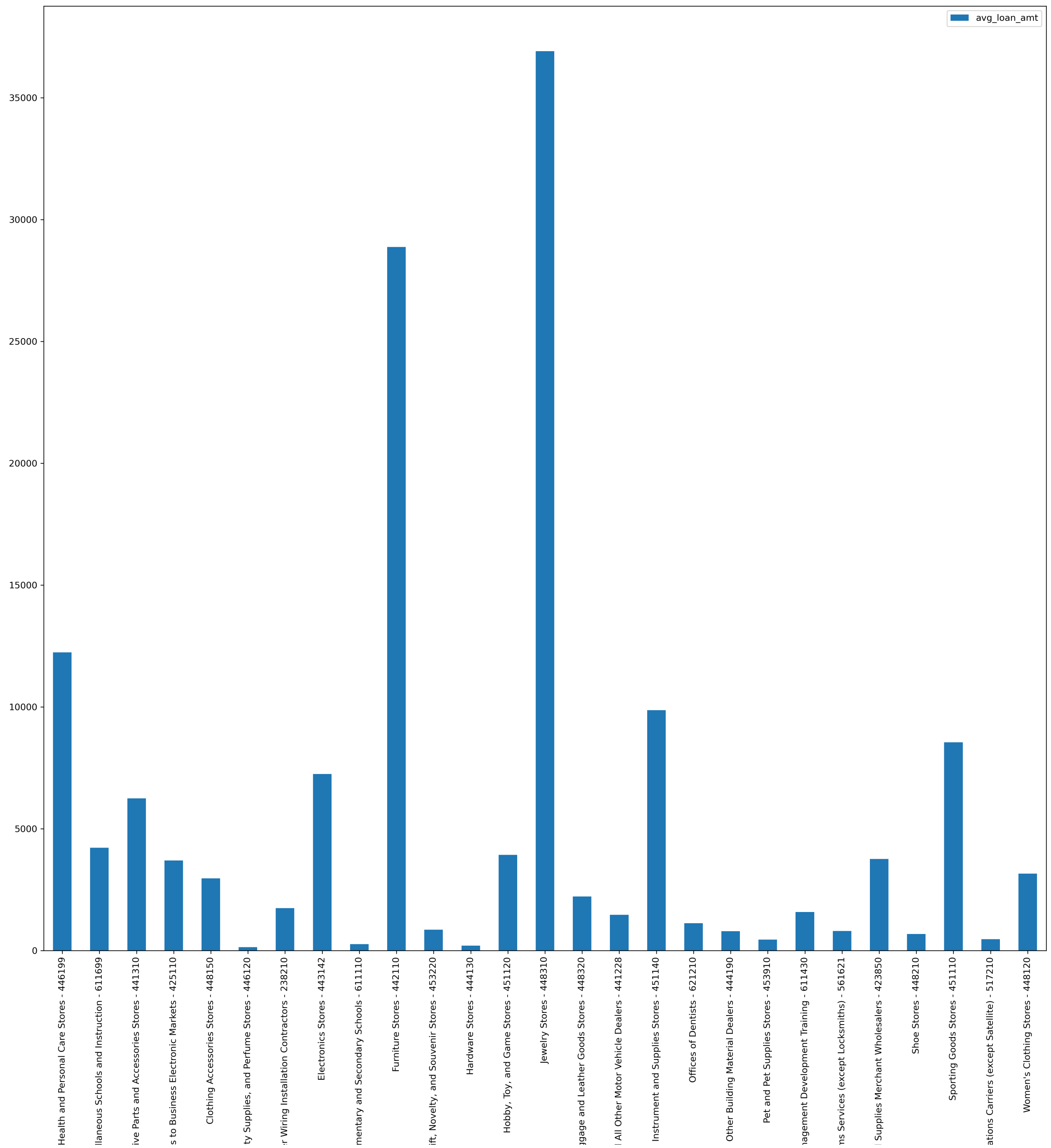
- The categories that had the highest average loan amount happened to be other, home finishing and jewelry.



Cont.

- The subcategories that had the highest average loan amount happened to be jewelry and furniture stores.




avg_loan_amt
35000
30000
25000
20000
15000
10000
5000
0
Health and Personal Care Stores - 446199
llaneous Schools and Instruction - 611699
ive Parts and Accessories Stores - 441310
s to Business Electronic Markets - 425110
Clothing Accessories Stores - 448150
ty Supplies, and Perfume Stores - 446120
r Wiring Installation Contractors - 238210
Electronics Stores - 443142
mentary and Secondary Schools - 611110
Furniture Stores - 442110
ift, Novelty, and Souvenir Stores - 453220
Hardware Stores - 444130
Hobby, Toy, and Game Stores - 451120
Jewelry Stores - 448310
ggage and Leather Goods Stores - 448320
All Other Motor Vehicle Dealers - 441228
Instrument and Supplies Stores - 451140
Offices of Dentists - 621210
Other Building Material Dealers - 444190
Pet and Pet Supplies Stores - 453910
agement Development Training - 611430
ns Services (except Locksmiths) - 561621
Supplies Merchant Wholesalers - 423850
Shoe Stores - 448210
Sporting Goods Stores - 451110
ations Carriers (except Satellite) - 517210
Women's Clothing Stores - 448120

Would you find the top 3 merchants with the most unit profit in each category?

- Here we have the top 3 merchants with the most unit profit for each subcategory. The way I did it was by using actual repayment percent, because if a company is paying back their loan, it is most likely because they are making profit.

	subcategory	actual_repayment_pct	name	merchant_id
122	All Other Health and Personal Care Stores - 44...	1.000000	KVGQTVZ	46716
128	All Other Health and Personal Care Stores - 44...	1.000000	UGGNLVQ	35999
86	All Other Health and Personal Care Stores - 44...	0.964568	AKGSYQP	83738
110	All Other Miscellaneous Schools and Instructio...	1.000000	NLQEXCA	88583
54	All Other Miscellaneous Schools and Instructio...	0.864338	YCBUYAB	53936
47	Automotive Parts and Accessories Stores - 441310	1.000000	PFZGFIM	75318
59	Automotive Parts and Accessories Stores - 441310	1.000000	IAJPFED	57481
12	Automotive Parts and Accessories Stores - 441310	0.884997	XVORKCX	64462
84	Business to Business Electronic Markets - 425110	1.000000	KSIUHKI	59868
101	Business to Business Electronic Markets - 425110	1.000000	LOBKKYS	72148
23	Business to Business Electronic Markets - 425110	0.971311	NJALGUT	30018
62	Clothing Accessories Stores - 448150	1.000000	OKSOYWN	24289
108	Clothing Accessories Stores - 448150	1.000000	CXBIUSH	51182
112	Clothing Accessories Stores - 448150	1.000000	HSYSUHE	18450
26	Cosmetics, Beauty Supplies, and Perfume Stores...	0.775577	HIXLDEB	39521
63	Electrical Contractors and Other Wiring Instal...	1.000000	SYEUKNX	12288
64	Electronics Stores - 443142	1.000000	BUOLYBH	47147
87	Electronics Stores - 443142	1.000000	DWTAPWD	20127
90	Electronics Stores - 443142	1.000000	HRXYKTP	42009
6	Elementary and Secondary Schools - 611110	0.899819	FUGCFOU	25861
49	Furniture Stores - 442110	1.000000	AFRBXMS	20513
57	Furniture Stores - 442110	1.000000	MIDVPQW	63991
61	Furniture Stores - 442110	1.000000	LHAYYTR	23023
44	Gift, Novelty, and Souvenir Stores - 453220	0.925765	JSCXBZP	73892
115	Hardware Stores - 444130	1.000000	NKEQGHV	24885
133	Hobby, Toy, and Game Stores - 451120	1.000000	FPFNBEI	28124
80	Hobby, Toy, and Game Stores - 451120	0.840283	IFAYZZM	15203
34	Hobby, Toy, and Game Stores - 451120	0.739957	OHJWQFF	83349
38	Jewelry Stores - 448310	1.000000	HRFSLMI	21132
69	Jewelry Stores - 448310	1.000000	QPHJHAL	78128
71	Jewelry Stores - 448310	1.000000	LFNRSGK	23058
51	Luggage and Leather Goods Stores - 448320	1.000000	UTNOWGK	12532
22	Luggage and Leather Goods Stores - 448320	0.948255	IGZQGDK	57280
93	Motorcycle, ATV, and All Other Motor Vehicle D...	1.000000	BZMPTGQ	80644
41	Musical Instrument and Supplies Stores - 451140	1.000000	HFVROIC	61893
66	Musical Instrument and Supplies Stores - 451140	1.000000	RPOREOR	13846
70	Musical Instrument and Supplies Stores - 451140	1.000000	WFALGEP	22950
15	Offices of Dentists - 621210	0.952664	CEYIWDR	47974
107	Other Building Material Dealers - 444190	1.000000	JPENRIH	85533
79	Pet and Pet Supplies Stores - 453910	1.000000	OOTKZJI	75173
53	Professional and Management Development Traini...	0.871179	XHHLRBL	50584
55	Security Systems Services (except Locksmiths) ...	0.963901	SSGHPQF	97926
72	Security Systems Services (except Locksmiths) ...	0.930312	IUYBVRJ	31029
81	Service Establishment Equipment and Supplies M...	1.000000	CEMTRAE	57265
105	Service Establishment Equipment and Supplies M...	0.848091	SOAFTOD	93651
14	Shoe Stores - 448210	0.944777	EGMCPIA	37889
58	Sporting Goods Stores - 451110	1.000000	DOUCDAQ	97508
67	Sporting Goods Stores - 451110	1.000000	DHTNLEJ	62512
129	Sporting Goods Stores - 451110	1.000000	BGEADFE	44770
43	Wireless Telecommunications Carriers (except S...	0.813168	DYIDJLP	60893
88	Wireless Telecommunications Carriers (except S...	0.307185	UCXMYZC	42733
92	Women's Clothing Stores - 448120	1.000000	NDDVOCX	63702
2	Women's Clothing Stores - 448120	0.936782	LYMLCLF	86469
0	Women's Clothing Stores - 448120	0.927466	MYRIFGH	78986

Cont.

- Here we have the top 3 merchants with the most unit profit for each category. The way I did it was by using actual repayment percent, because if a company is paying back their loan, it is most likely because they are making profit.

	category	actual_repayment_pct	name	merchant_id
47	AUTO_PARTS	1.000000	PFZGFIM	75318
59	AUTO_PARTS	1.000000	IAJPFED	57481
12	AUTO_PARTS	0.884997	XVORKCX	64462
26	BEAUTY	0.775577	HIXLDEB	39521
41	CONSUMER_ELECTRONICS	1.000000	HFVROIC	61893
84	CONSUMER_ELECTRONICS	1.000000	KSIUHKI	59868
87	CONSUMER_ELECTRONICS	1.000000	DWTAPWD	20127
49	HOME_FURNISHINGS	1.000000	AFRBXMS	20513
57	HOME_FURNISHINGS	1.000000	MIDVPQW	63991
61	HOME_FURNISHINGS	1.000000	LHAYYTR	23023
38	JEWELRY	1.000000	HRFSLMI	21132
69	JEWELRY	1.000000	QPHJHAL	78128
74	JEWELRY	1.000000	EKBOCBN	34543
62	MENS_FASHION	1.000000	OKSOYWN	24289
58	OTHER	1.000000	DOUCDAQ	97508
66	OTHER	1.000000	RPOREOR	13846
67	OTHER	1.000000	DHTNLEJ	62512
110	PERSONAL_SERVICE	1.000000	NLQEXCA	88583
15	PERSONAL_SERVICE	0.952664	CEYIWDR	47974
51	WOMENS_FASHION	1.000000	UTNOWGK	12532
92	WOMENS_FASHION	1.000000	NDDVOCX	63702
108	WOMENS_FASHION	1.000000	CXBIUSH	51182

Based on the analysis, in which areas would you increase or decrease volume?

- Based on the analysis, an area I would decrease volume in the Jewelry store subcategory. Due to the fact, they have one of the highest number of loan amount and one of longest loan terms. Having a very long term can be risky, because they are taking longer to pay back, mainly when it is a large loan. I would increase beauty, because they have one of the lowest loan volume and term length. Not only that, but they tend to have one of the lowest APRs.

What information can Company XYZ gather to further evaluate their merchants' profit with Affirm? What types of analysis, evaluation, or diligence should Company XYZ do?

- The information Company XYZ can gather to further evaluate their merchants' profit with Affirm is annual income, collateral, savings, and demographics of the merchants. Annual income is beneficial due to the fact, that you can see how much a merchant is making. This allows you to see how much they are selling. This can indicate if a merchant has the income to profit and pay their loans. Collateral is good to have as security in place, this can further be used to evaluate risk levels towards a merchant. While also seeing savings, this can allow Affirm to see how much they have saved in case of an emergency and if they have the funds to stay afloat. Demographics is another piece of information to have in mind because you can see who is coming to buy, how much they are buying, and willing to pay back when they take out a loan with Affirm via the merchant. Not only that but you can see their income to know if they can truly afford a purchase in the end.