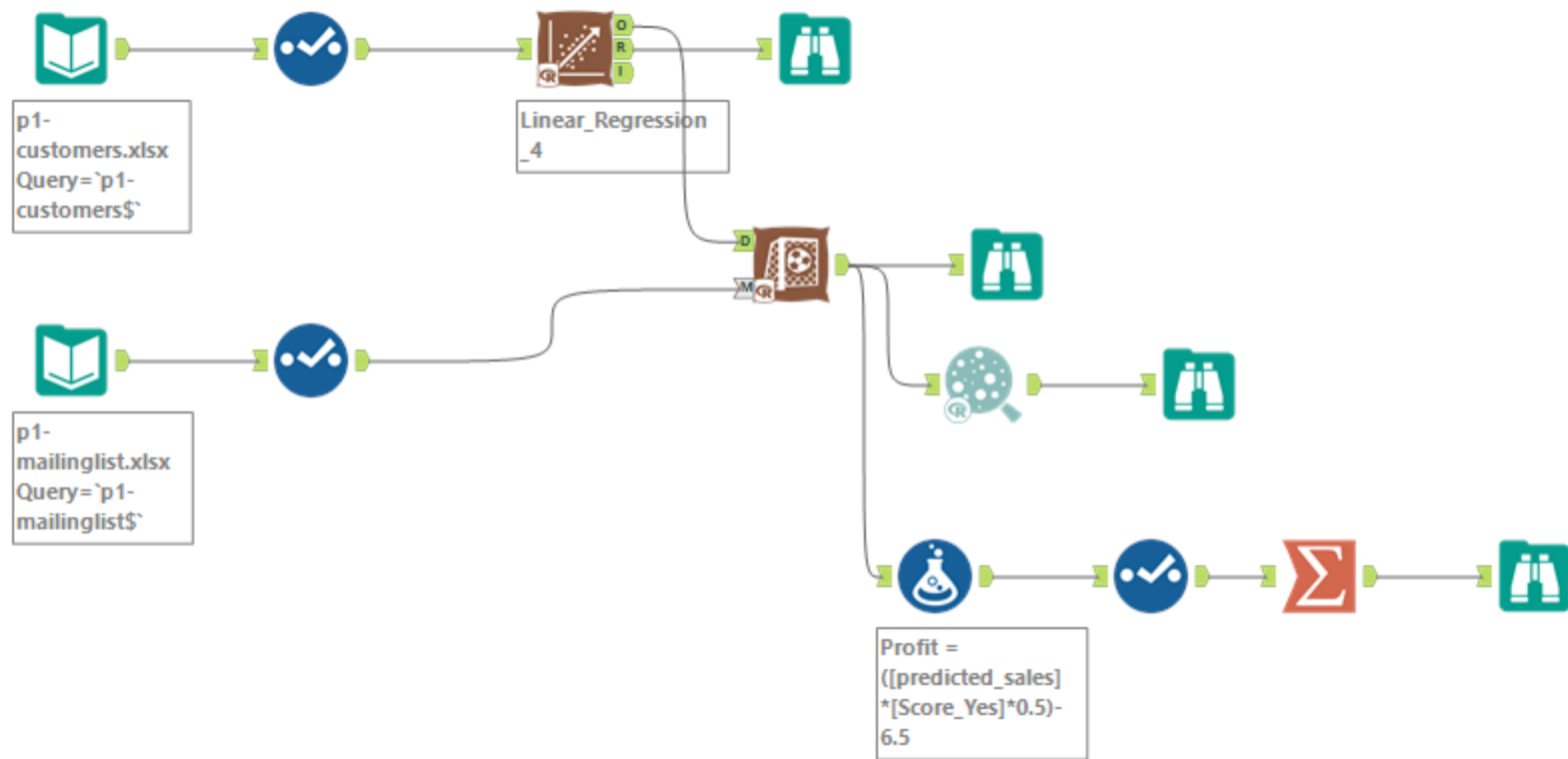
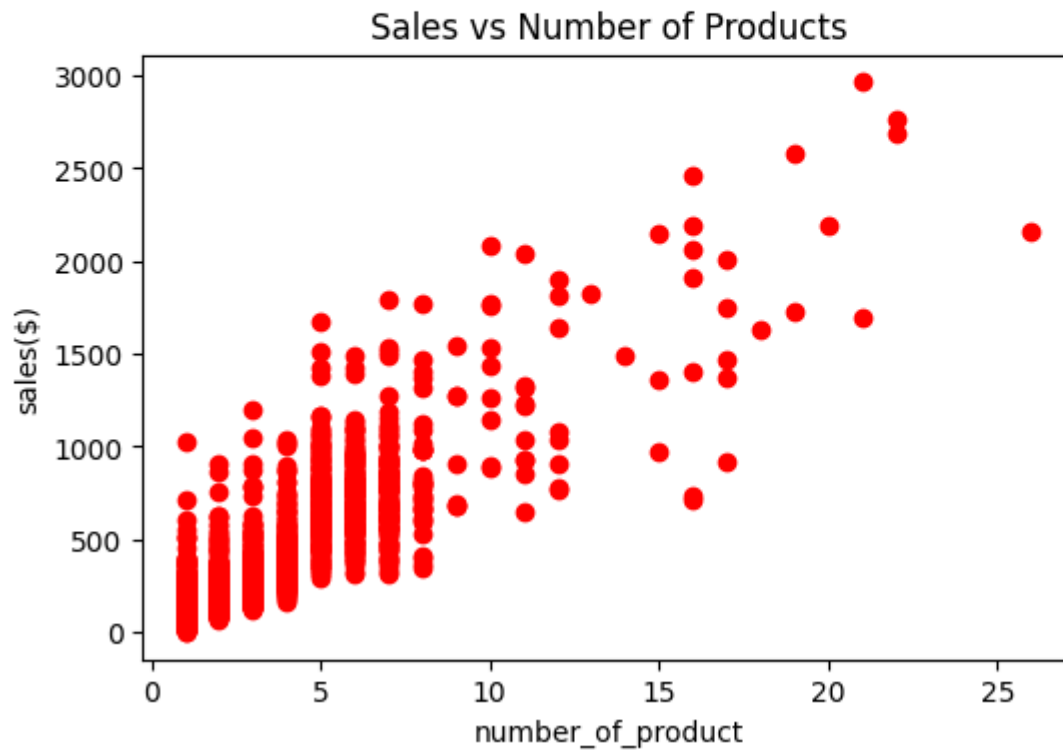


Predicting Catalog Demand

Alteryx - Excel - Python



```
fig, ax = plt.subplots()
ax.scatter(training_data["Avg_Num_Products_Purchased"], training_data["Avg_Sale_Amount"], color = "r")
ax.set_xlabel("number_of_product")
ax.set_ylabel("sales($)")
ax.set_title("Sales vs Number of Products")
plt.show()
```



The scatter plot below shows a positive linear relationship between the continuous predictor variable and the target variable. The "number_of_product" increases as the "sales" increases.

Table Report

1 of 1 Fields 1 of 1 Fields Records 1 to 10

Record	Report																														
1	Report for Linear Model Linear_Regression_4																														
2	<i>Basic Summary</i>																														
3	Call: lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)																														
4	Residuals:																														
5	<table><tr><th>Min</th><th>1Q</th><th>Median</th><th>3Q</th><th>Max</th></tr><tr><td>-663.8</td><td>-67.3</td><td>-1.9</td><td>70.7</td><td>971.7</td></tr></table>	Min	1Q	Median	3Q	Max	-663.8	-67.3	-1.9	70.7	971.7																				
Min	1Q	Median	3Q	Max																											
-663.8	-67.3	-1.9	70.7	971.7																											
6	Coefficients:																														
7	<table><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr><tr><td>(Intercept)</td><td>303.46</td><td>10.576</td><td>28.69</td><td>< 2.2e-16 ***</td></tr><tr><td>Customer_SegmentLoyalty Club Only</td><td>-149.36</td><td>8.973</td><td>-16.65</td><td>< 2.2e-16 ***</td></tr><tr><td>Customer_SegmentLoyalty Club and Credit Card</td><td>281.84</td><td>11.910</td><td>23.66</td><td>< 2.2e-16 ***</td></tr><tr><td>Customer_SegmentStore Mailing List</td><td>-245.42</td><td>9.768</td><td>-25.13</td><td>< 2.2e-16 ***</td></tr><tr><td>Avg_Num_Products_Purchased</td><td>66.98</td><td>1.515</td><td>44.21</td><td>< 2.2e-16 ***</td></tr></table> <p>Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	303.46	10.576	28.69	< 2.2e-16 ***	Customer_SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16 ***	Customer_SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16 ***	Customer_SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16 ***	Avg_Num_Products_Purchased	66.98	1.515	44.21	< 2.2e-16 ***
	Estimate	Std. Error	t value	Pr(> t)																											
(Intercept)	303.46	10.576	28.69	< 2.2e-16 ***																											
Customer_SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16 ***																											
Customer_SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16 ***																											
Customer_SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16 ***																											
Avg_Num_Products_Purchased	66.98	1.515	44.21	< 2.2e-16 ***																											
8	Residual standard error: 137.48 on 2370 degrees of freedom Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366 F-statistic: 3040 on 4 and 2370 degrees of freedom (DF), p-value < 2.2e-16																														
9	<i>Type II ANOVA Analysis</i>																														
10	Response: Avg_Sale_Amount <table><tr><th></th><th>Sum Sq</th><th>DF</th><th>F value</th><th>Pr(>F)</th></tr><tr><td>Customer_Segment</td><td>28715078.96</td><td>3</td><td>506.4</td><td>< 2.2e-16 ***</td></tr><tr><td>Avg_Num_Products_Purchased</td><td>36939582.5</td><td>1</td><td>1954.31</td><td>< 2.2e-16 ***</td></tr><tr><td>Residuals</td><td>44796869.07</td><td>2370</td><td></td><td></td></tr></table> <p>Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p>		Sum Sq	DF	F value	Pr(>F)	Customer_Segment	28715078.96	3	506.4	< 2.2e-16 ***	Avg_Num_Products_Purchased	36939582.5	1	1954.31	< 2.2e-16 ***	Residuals	44796869.07	2370												
	Sum Sq	DF	F value	Pr(>F)																											
Customer_Segment	28715078.96	3	506.4	< 2.2e-16 ***																											
Avg_Num_Products_Purchased	36939582.5	1	1954.31	< 2.2e-16 ***																											
Residuals	44796869.07	2370																													

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	490.9834	123.042	3.99038	7e-05	****
Customer_SegmentLoyalty Club Only	-149.8780	9.020	-16.61610	< 2.2e-16	****
Customer_SegmentLoyalty Club and Credit Card	282.9160	11.971	23.63433	< 2.2e-16	****
Customer_SegmentStore Mailing List	-245.5024	9.846	-24.93384	< 2.2e-16	****
CityAurora	-19.9592	11.097	-1.79862	0.07221	.
CityBoulder	-40.3075	80.127	-0.50304	0.61498	
CityBrighton	-70.4394	97.670	-0.72120	0.47086	
CityBroomfield	-3.8360	15.142	-0.25334	0.80003	
CityCastle Pines	-91.0506	97.760	-0.93137	0.35176	
CityCentennial	-9.9043	18.180	-0.54479	0.58595	
CityCommerce City	-35.5183	44.500	-0.79816	0.42486	
CityDenver	-0.4258	10.561	-0.04032	0.96784	
CityEdgewater	29.6501	40.657	0.72928	0.46591	
CityEnglewood	5.0515	20.760	0.24333	0.80777	
CityGolden	-12.8415	32.755	-0.39205	0.69506	
CityGreenwood Village	-50.0735	38.045	-1.31616	0.18825	
CityHenderson	-284.9016	138.017	-2.06425	0.0391	*
CityHighlands Ranch	-27.7328	30.457	-0.91055	0.36263	
CityLafayette	-47.4442	62.200	-0.76277	0.44568	
CityLakewood	-7.9770	12.872	-0.61974	0.53549	
CityLittleton	-28.8015	18.991	-1.51663	0.1295	
CityLone Tree	77.5542	137.939	0.56223	0.57401	
CityLouisville	-28.6488	69.348	-0.41312	0.67956	
CityMorrison	-17.2654	52.851	-0.32668	0.74394	
CityNorthglenn	-15.2884	29.428	-0.51952	0.60345	
CityParker	-6.0407	28.212	-0.21411	0.83048	
CitySuperior	-53.5206	46.728	-1.14535	0.25218	
CityThornton	28.5171	24.843	1.14787	0.25114	
CityWestminster	-6.8925	17.305	-0.39829	0.69045	
CityWheat Ridge	7.1755	20.687	0.34685	0.72873	
Store_Number	-1.6275	1.148	-1.41823	0.15626	
Avg_Num_Products_Purchased	67.1063	1.527	43.93513	< 2.2e-16	****
X_ _Years_as_Customer	-2.3582	1.232	-1.91353	0.0558	.

Type II ANOVA Analysis

Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)	
Customer_Segment	28404574.79	3	500.78	< 2.2e-16	***
City	423523.12	26	0.86	0.66618	✗
Store_Number	38029.04	1	2.01	0.15626	✗
Avg_Num_Products_Purchased	36495947.85	1	1930.3	< 2.2e-16	***
X._Years_as_Customer	69229.79	1	3.66	0.0558	✗
Residuals	44280006.63	2342			

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	303.46	10.576	28.69	< 2.2e-16	***
Customer_SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16	***
Customer_SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16	***
Customer_SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16	***
Avg_Num_Products_Purchased	66.98	1.515	44.21	< 2.2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 137.48 on 2370 degrees of freedom

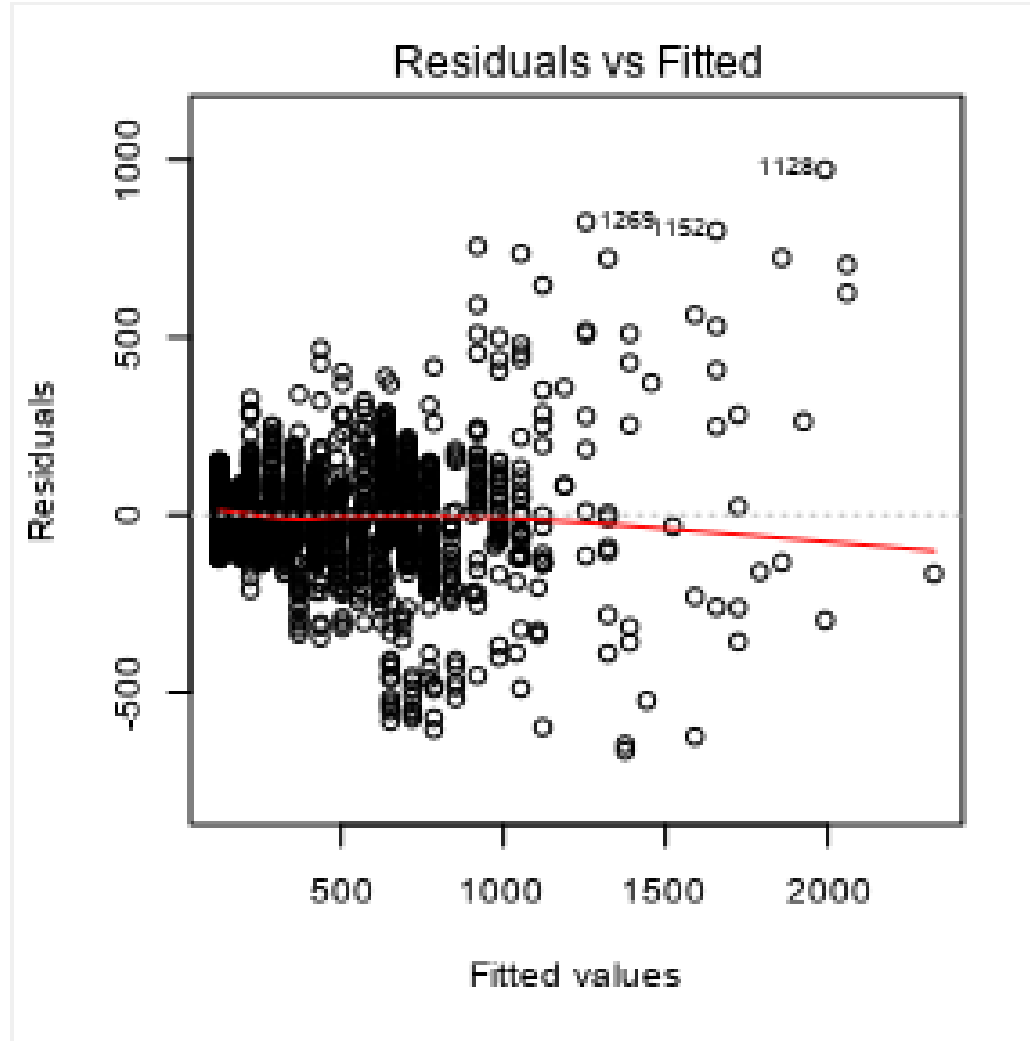
Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366

F-statistic: 3040 on 4 and 2370 degrees of freedom (DF), p-value < 2.2e-16

Type II ANOVA Analysis

Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)	
Customer_Segment	28715078.96	3	506.4	< 2.2e-16	***
Avg_Num_Products_Purchased	36939582.5	1	1954.31	< 2.2e-16	***
Residuals	44796869.07	2370			



Schema Information

Field #1

Record	Sum_Profit
1	21,987.435687

Schema Information

Field #1

Name: Customer_Segment

Type: V_WString

Size: 1073741823

Source: R-DATA:

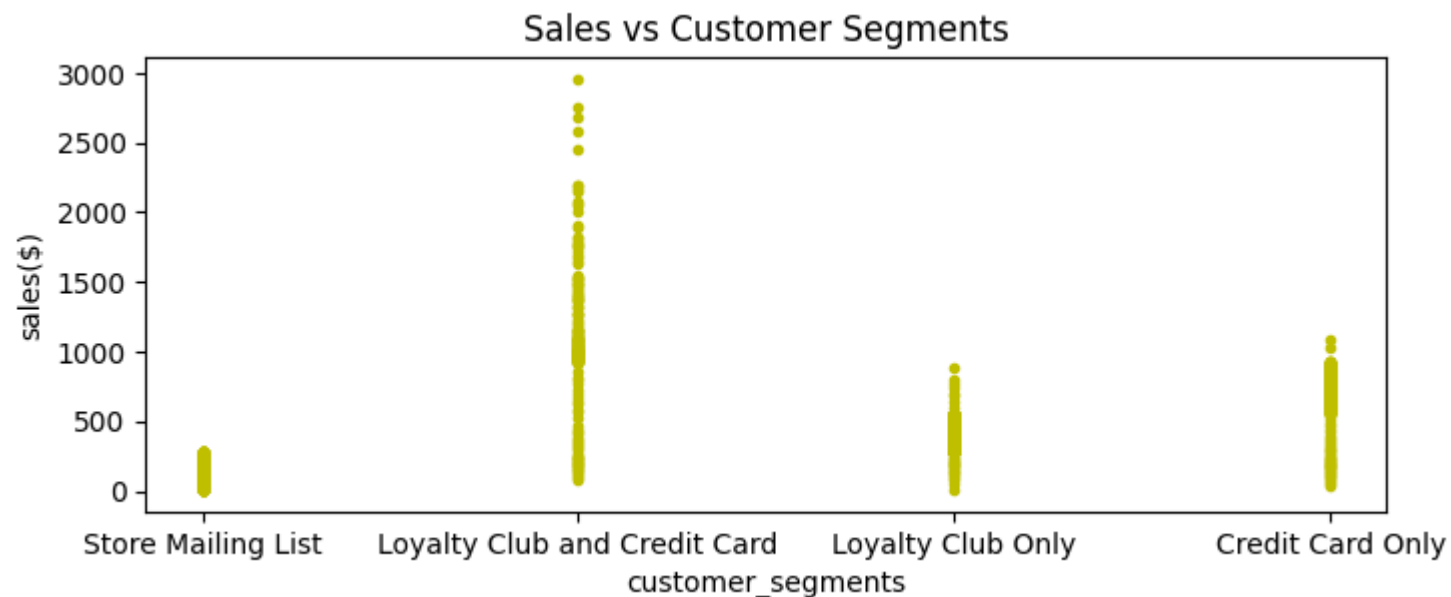
Description:

<

Record	Customer_Segment	Avg_Num_Products_Purchased	Score_Yes	predicted_sales	Profit
1	Loyalty Club Only	3	0.305036	355.036364	47.649402
2	Loyalty Club and Credit Card	6	0.472725	987.159466	226.827251
3	Loyalty Club Only	7	0.578882	622.941184	173.804672
4	Loyalty Club Only	2	0.305138	288.060159	37.449023
5	Loyalty Club Only	4	0.387706	422.012569	75.308372
6	Credit Card Only	7	0.267278	772.296906	96.709099
7	Loyalty Club and Credit Card	4	0.221739	853.207056	88.09485
8	Credit Card Only	6	0.193447	705.320701	61.721139
9	Credit Card Only	6	0.250658	705.320701	81.897002
10	Loyalty Club Only	4	0.264523	422.012569	49.316047
11	Store Mailing List	2	0.190541	191.998137	11.791797
12	Loyalty Club Only	7	0.191545	622.941184	53.160605
13	Loyalty Club Only	4	0.212284	422.012569	38.293279
14	Credit Card Only	5	0.277962	638.344496	82.217787
15	Credit Card Only	7	0.26971	772.296906	97.647926
16	Loyalty Club Only	3	0.238436	355.036364	35.82672


```
| grouped = training_data.groupby('Customer_Segment')['Avg_Sale_Amount'].sum()  
grouped_sort = grouped.sort_values(ascending=False)  
grouped_sort
```

```
Customer_Segment  
Credit Card Only      337243.40  
Loyalty Club Only      229476.56  
Loyalty Club and Credit Card  208386.94  
Store Mailing List     174356.57  
Name: Avg_Sale_Amount, dtype: float64
```



```
grouped = training_data.groupby('City')['Avg_Sale_Amount'].sum()
grouped_sort = grouped.sort_values(ascending=False)
grouped_sort
```

City	
Denver	303478.07
Aurora	199586.67
Arvada	95363.39
Lakewood	90649.20
Broomfield	49186.92
Westminster	33618.94
Centennial	29037.10
Littleton	27097.90
Wheat Ridge	23217.91
Englewood	21071.73
Thornton	13799.98
Parker	9045.74
Highlands Ranch	8992.79
Northglenn	8562.10
Golden	7093.98
Commerce City	6828.15
Edgewater	5555.54
Greenwood Village	4564.68
Morrison	3588.86
Superior	2745.18
Lafayette	1720.66
Boulder	1620.57
Louisville	1393.87
Lone Tree	854.87
Castle Pines	384.42
Henderson	214.39
Brighton	189.86

Name: Avg_Sale_Amount, dtype: float64

```
total = training_data['Avg_Sale_Amount'].sum() # total = $949463.47
Denver_ = round(303478.07/total*100,0)
print(Denver_, '%')
```

32.0 %

```
Brighton_ = round(189.86/total*100,2)
print(Brighton_, '%')
```

0.02 %

```
fig, ax = plt.subplots()
ax.scatter(training_data["#_Years_as_Customer"], training_data["Avg_Sale_Amount"], color = "b", marker = '.')
ax.set_xlabel("years_as_customer ")
ax.set_ylabel("sales($)")
ax.set_title("Sales vs Number of Years as Customers")
plt.show()
```

