OLIST:

linear regression models and customer lifetime value

VADIM ACOSTA

The setup

- **Olist** - integrates sellers and marketplaces to create a commercial platform across multiple channels
- **Goal** – create a **C**ustomer **L**ifetime **V**alue model to predict and understand customer spending

The Tools

- Excel
- Tableau
- Python, Jupyter Notebook, Pandas *et al.*

The Plan

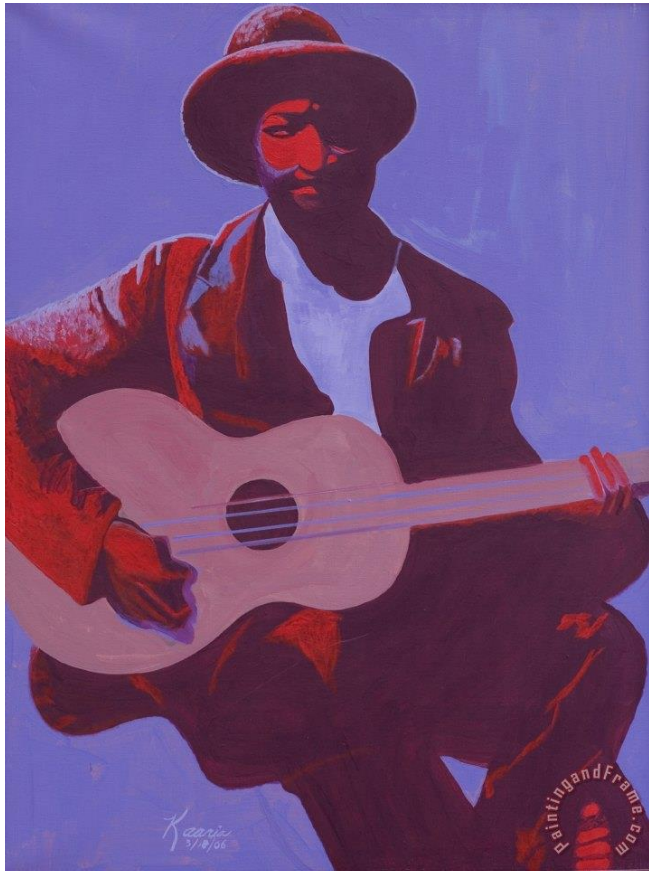
- Consolidate relevant data
- Remove nulls
- Remove irrelevant columns
- Group data by **Unique Customer ID**
- Run Regressions to model CLV

The Linear Regression Blues

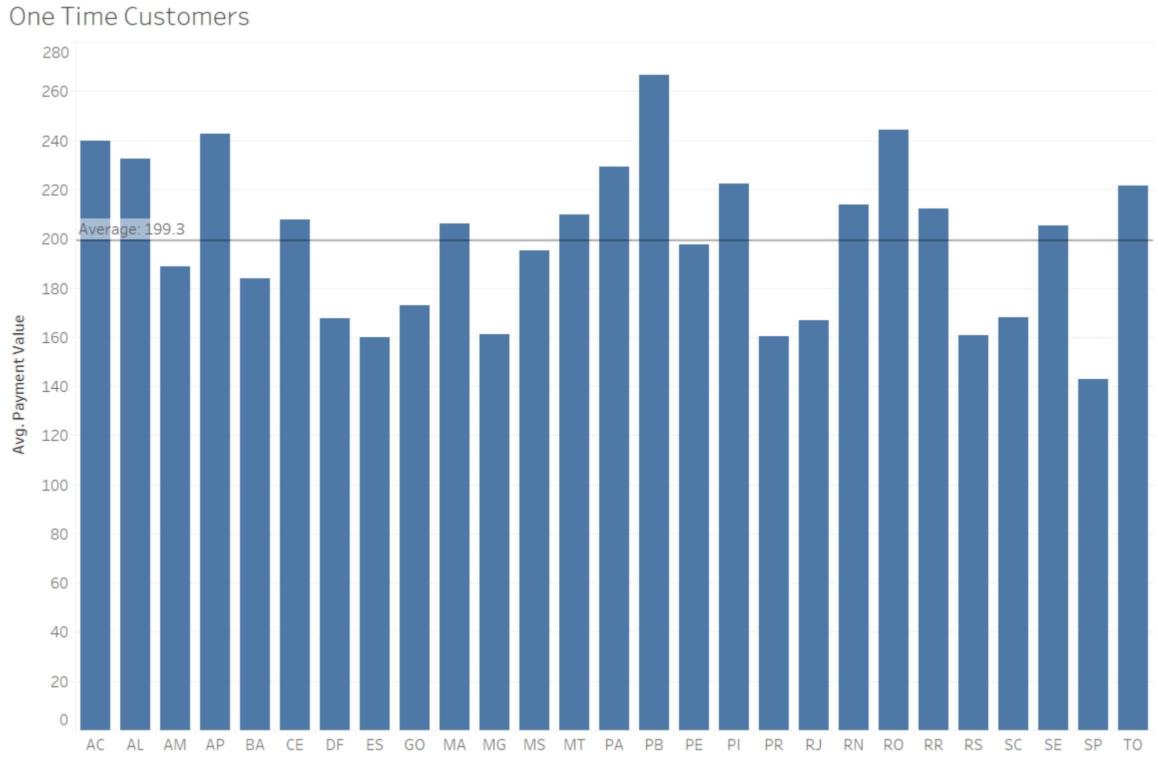
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	83.9834	1.635	51.370	0.000	80.779	87.188
payment_installments	26.0526	0.238	109.326	0.000	25.586	26.520
customer_zip_code_prefix	0.0003	2.14e-05	12.913	0.000	0.000	0.000
payment_sequential	-13.7137	0.979	-14.011	0.000	-15.632	-11.795

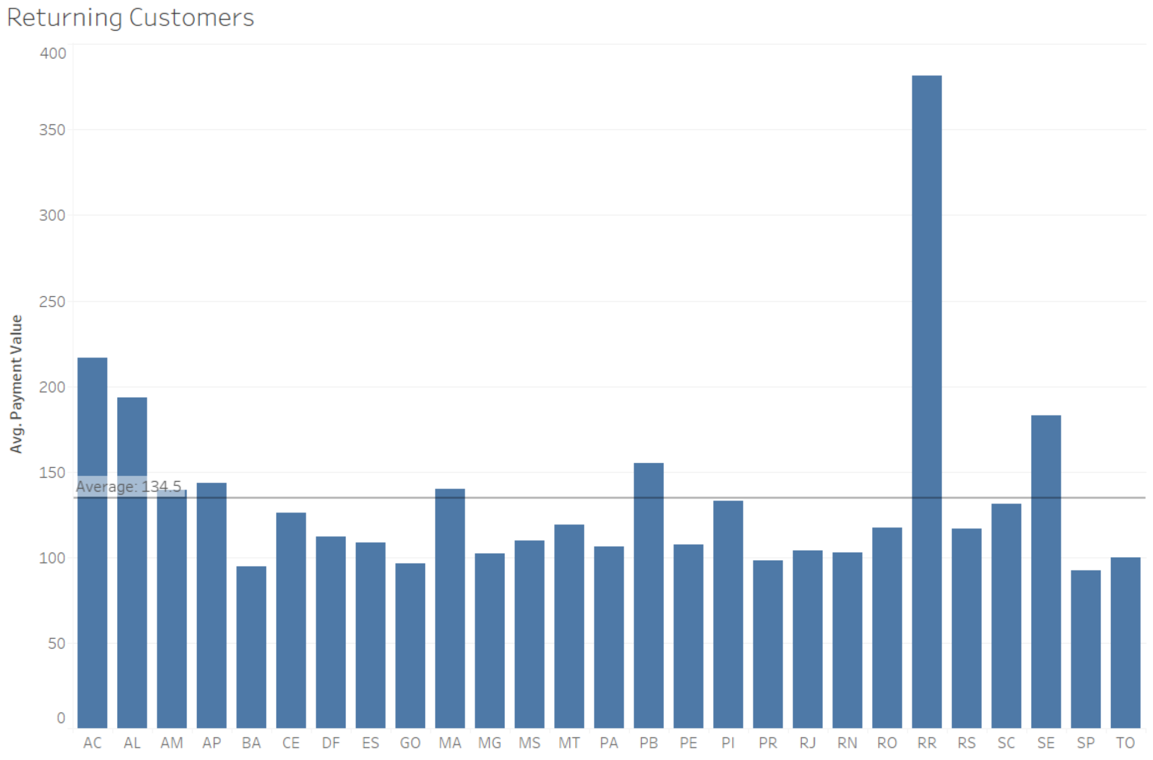
Dep. Variable:	payment_value	**R-squared:**	0.113
Model:	OLS	**Adj. R-squared:**	0.113
Method:	Least Squares	**F-statistic:**	4299.
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	0.00
Time:	12:17:40	**Log-Likelihood:**	-6.7766e+05
No. Observations:	100739	**AIC:**	1.355e+06
Df Residuals:	100735	**BIC:**	1.355e+06
Df Model:	3		
Covariance Type:	nonrobust		

- **P-values** are low (0.000), implying that each of the independent variables exert meaningful influence on the dependent variable (payment value)
- **R^2** values are also low (0.113), implying that the model doesn't fit the data very well



Return Customers Spend Less Money Per Purchase





The New Plan

- Separate customers that made only one order from those that made multiple orders
- Use “payments made” as an independent variable to model lifetime value

Using Payment Value to predict total value

abridged_join_data_all_customers_final - Excel

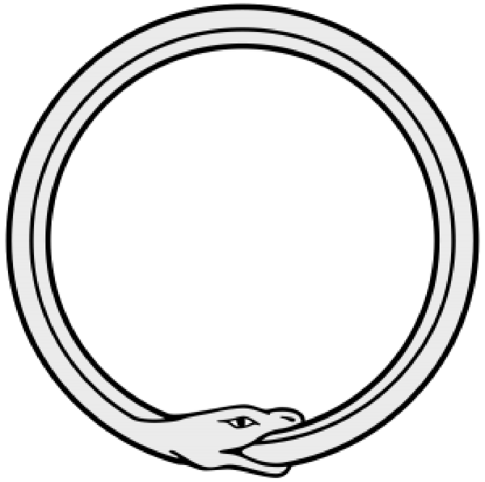
P1 total_value

	D	J	K	L	M	N	O	P
1	customer_unique_id	payment_installments	payment_value	calculation_1	calculation_2	calculation_3	calculation_4	total_value
2	7c396fd4830fd04220f754e42b4e5bff	1	18.12	18.12	0	0	0	38.71
3	7c396fd4830fd04220f754e42b4e5bff	1	2	20.12	0	0	38.71	38.71
4	7c396fd4830fd04220f754e42b4e5bff	1	18.59	38.71	38.71	38.71	38.71	38.71
5	af07308b275d755c9edb36a90c618231	1	141.46	141.46	141.46	141.46	141.46	141.46
6	3a653a41f6f9fc3d2a113cf8398680e8	3	179.12	179.12	179.12	179.12	179.12	179.12
7	7c142cf63193a1473d2e66489a9ae977	1	72.2	72.2	72.2	72.2	72.2	72.2
8	72632f0f9dd73dfee390c9b22eb56dd6	1	28.62	28.62	28.62	28.62	28.62	28.62
9	80bb27c7c16e8f973207a5086ab329e2	6	175.26	175.26	175.26	175.26	175.26	175.26
10	36edbb3fb164b1f16485364b6fb04c73	1	65.95	65.95	65.95	65.95	65.95	65.95
11	932afa1e708222e5821dac9cd5db4cae	3	75.16	75.16	75.16	75.16	75.16	75.16
12	39382392765b6dc74812866ee5ee92a7	1	35.95	35.95	35.95	35.95	35.95	35.95
13	299905e3934e9e181bfb2e164dd4b4f8	1	161.42	161.42	0	0	169.76	169.76
14	299905e3934e9e181bfb2e164dd4b4f8	1	8.34	169.76	169.76	169.76	169.76	169.76
15	f2a85dec752b8517b5e58a06ff3cd937	1	259.06	259.06	259.06	259.06	259.06	259.06

Modeling lifetime value based on payments made is meaningless when the customer only makes **one** purchase

Dep. Variable:	total_value	**R-squared:**	1.000
Model:	OLS	**Adj. R-squared:**	1.000
Method:	Least Squares	**F-statistic:**	6.253e+27
Date:	Mon, 09 Dec 2019	**Prob (F-statistic):**	0.00
Time:	09:29:44	**Log-Likelihood:**	1.8661e+06
No. Observations:	88058	**AIC:**	-3.732e+06
Df Residuals:	88027	**BIC:**	-3.732e+06
Df Model:	30		
Covariance Type:	nonrobust		

If the customer **only** makes **one** payment in his lifetime then his lifetime value will have a **100% match** with his one payment



We can use linear regression to make fairly accurate predictions about people who make multiple payments

Dep. Variable:	total_value	**R-squared:**	0.707
Model:	OLS	**Adj. R-squared:**	0.706
Method:	Least Squares	**F-statistic:**	573.5
Date:	Mon, 09 Dec 2019	**Prob (F-statistic):**	0.00
Time:	09:28:40	**Log-Likelihood:**	-79174.
No. Observations:	12901	**AIC:**	1.585e+05
Df Residuals:	12846	**BIC:**	1.589e+05
Df Model:	54		
Covariance Type:	nonrobust		

Conclusion

- We **can't** use linear regression to predict anything about people who only made one purchase, which is more than **80%** of our customers

- We **can** use linear regression to make fairly accurate predictions about our **best** customers

Recommendations

- Encourage customers to use installments and vouchers to pay less money but more frequently
- Send marketing materials to customers that are approaching their final installments to get them on a new round of payments

Thank you!

NYC Tech Talent Pipeline

galvanize