Walmart

CAPSTONE PROJECT

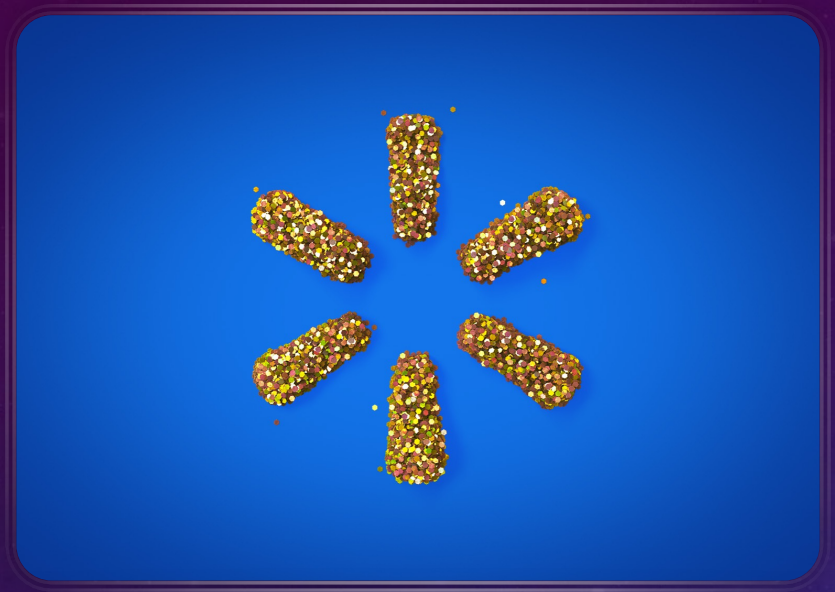
Author: Christopher Freyre

CONTENTS

- Business Problem
- Data understanding
- Data cleaning
- Model and Validation

BUSINESS PROBLEM

Walmart has been running out of stock during busy periods recently and are looking for a way to predict future sales in order to maintain appropriate levels of stock.



	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106
...	...	...	...	...	...	...	...	...
6430	45	28-09-2012	713173.95	0	64.88	3.997	192.013558	8.684
6431	45	05-10-2012	733455.07	0	64.89	3.985	192.170412	8.667
6432	45	12-10-2012	734464.36	0	54.47	4.000	192.327265	8.667
6433	45	19-10-2012	718125.53	0	56.47	3.969	192.330854	8.667
6434	45	26-10-2012	760281.43	0	58.85	3.882	192.308899	8.667

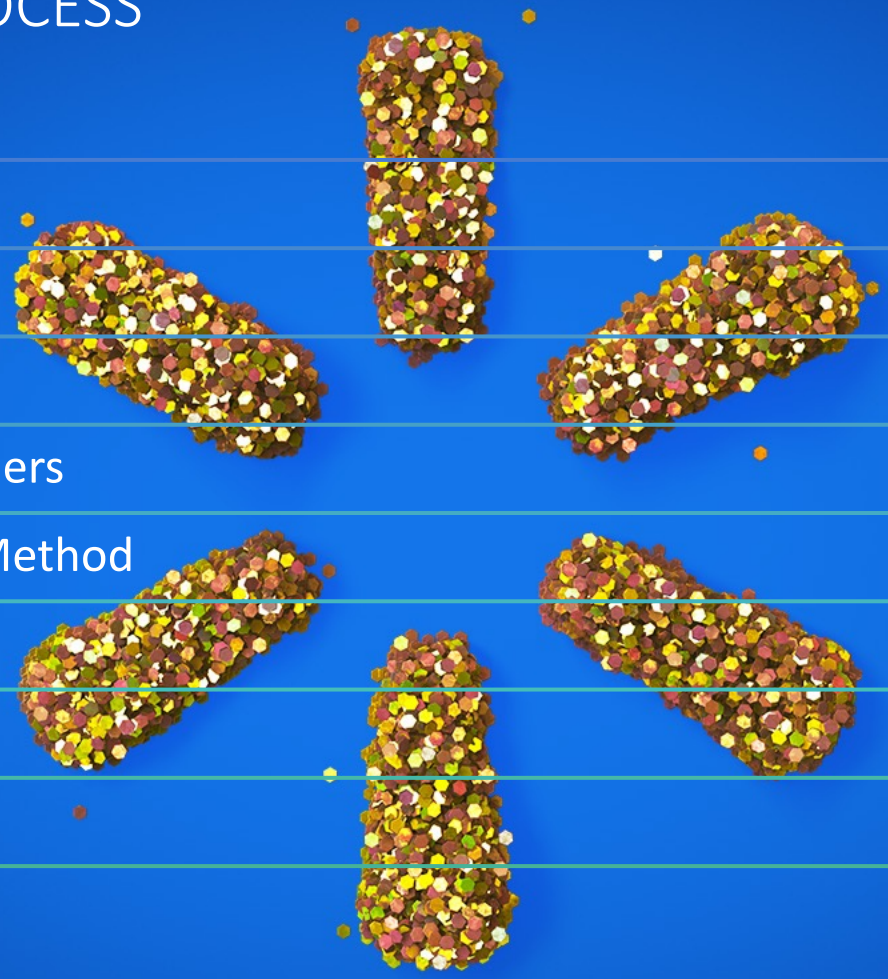
6435 rows × 8 columns

DATA UNDERSTANDING

- Stores - 45
- Date (2010-2012)
- Weekly Sales
- Holiday Sales
- Temperature - °F
- Fuel Price
- CPI – Consumer price index
- Unemployment

DATA CLEANING PROCESS

- Missing values
- Unique values
- Transformation Datetime
- Checking and removing Outliers
- IQR Trimming and Capping Method
- Log Transformation
- Scaling
- Modeling
- Validation



MODEL AND VALIDATION

- Improvement of models
- Performing creating of dummies variables for categorical values
- Performing Logarithmic transformation
- Scaling Data values
- Validation T-test
- Cross validation method

VALIDATION

R-square value:
0.805

Coeff values:
Q1<Q2<Q3<Q4
Small<Medium<Large
Days<Holiday Days

T-test scores:
Train MSE = 0.0377
Test MSE = 0.0387

Cross Validation:
MSE = 0.0874

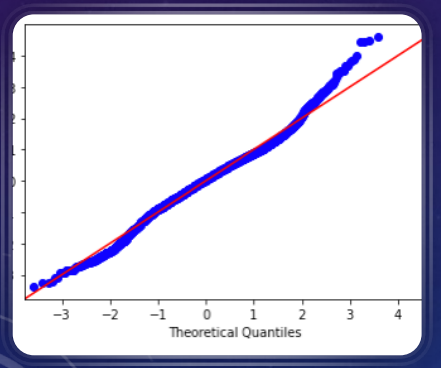
OLS Regression Results

Dep. Variable:	Weekly_Sales	**R-squared:**	0.805
Model:	OLS	**Adj. R-squared:**	0.805

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	7.2791	0.051	142.298	0.000	7.179	7.379
Store_type_small	1.7838	0.018	99.278	0.000	1.749	1.819
Store_type_medium	2.4584	0.017	143.228	0.000	2.425	2.492
Store_type_large	3.0369	0.018	168.384	0.000	3.001	3.072
Holiday_Flag_0	3.6386	0.026	137.701	0.000	3.587	3.690
Holiday_Flag_1	3.6405	0.027	136.110	0.000	3.588	3.693
Temperature	-0.1126	0.011	-10.471	0.000	-0.134	-0.091
Fuel_Price	0.0043	0.025	0.170	0.865	-0.045	0.054
CPI	-0.1524	0.016	-9.753	0.000	-0.183	-0.122
Unemployment	-0.1048	0.021	-5.058	0.000	-0.145	-0.064
Quarter_Q1	1.7719	0.014	126.485	0.000	1.744	1.799
Quarter_Q2	1.8248	0.014	126.958	0.000	1.797	1.853
Quarter_Q3	1.8329	0.014	126.421	0.000	1.805	1.861
Quarter_Q4	1.8494	0.014	132.452	0.000	1.822	1.877

CONCLUSION

- Model_q3 provided the best fit for linear regression with an R-squared value of 0.805. Meaning that it represents 80.5% of the data.
- Validating it with T-test, provided a Mean Squared Error value: 0.0377.
- Cross Validation result: 0.087
- Important Coeff:
  - Quarters
  - Store Types
  - Holiday Flag



Quarter	Store_type	count	mean	std	min	25%	50%	75%	max
Q1	small	486	484,045.07	159,228.97	209,986.25	367,819.65	486,527.82	561,206.07	1,376,520.10
	medium	416	931,800.51	197,582.81	540,922.94	786,004.20	912,004.09	1,065,870.84	1,617,025.41
	large	461	1,655,897.49	312,866.43	1,057,290.41	1,379,473.03	1,632,616.09	1,910,092.37	2,495,630.51
Q2	small	577	495,059.56	169,885.86	234,218.03	369,350.60	492,364.77	573,498.64	1,500,863.54
	medium	494	973,090.14	196,028.14	596,218.24	814,046.84	964,040.24	1,100,453.11	1,571,158.56
	large	544	1,696,227.06	303,212.52	1,194,334.65	1,422,773.59	1,654,382.89	1,945,633.90	2,623,469.95
Q3	small	588	507,347.77	192,148.00	224,031.19	367,167.22	501,098.52	583,112.69	1,469,693.99
	medium	481	960,509.55	208,877.79	558,794.63	802,583.89	937,420.65	1,095,889.22	1,841,173.60
	large	522	1,654,245.53	289,681.24	1,169,413.27	1,391,943.32	1,614,016.62	1,886,116.50	2,546,123.78
Q4	small	501	530,257.27	218,226.15	213,538.32	389,540.62	509,647.25	598,437.98	1,648,829.18
	medium	422	1,061,263.70	316,811.67	576,332.05	847,406.27	997,685.55	1,184,727.22	2,587,953.32
	large	425	1,733,915.83	347,651.96	1,169,831.38	1,425,078.59	1,707,298.14	1,984,768.34	2,685,351.81

RECOMMENDATIONS

Using my final model, I advise Walmart Stores:

- Demand will increase quarterly and will be significantly higher during the last quarter of the year.
- Categorising stores into small, medium and large will help prepare them for future demand and stock supply
- Demand amount will depend on the store size, the model tells us that the larger the store, the higher the demand.
- Be prepared for holiday days as they are a factor that will increase demand.

THANK YOU

Email: Christopher.Freyre@gmail.com
Github: @Chrisfreyre
Linkedin: https://www.linkedin.com/in/christopher-freyre-56760b136