CAPSTONE PROJECT 4

Introduction

- A sales analyst trying to predict Lifetime Post Total Reach with various factors
- To provide insights for the sales team on how to further improve sales

Source

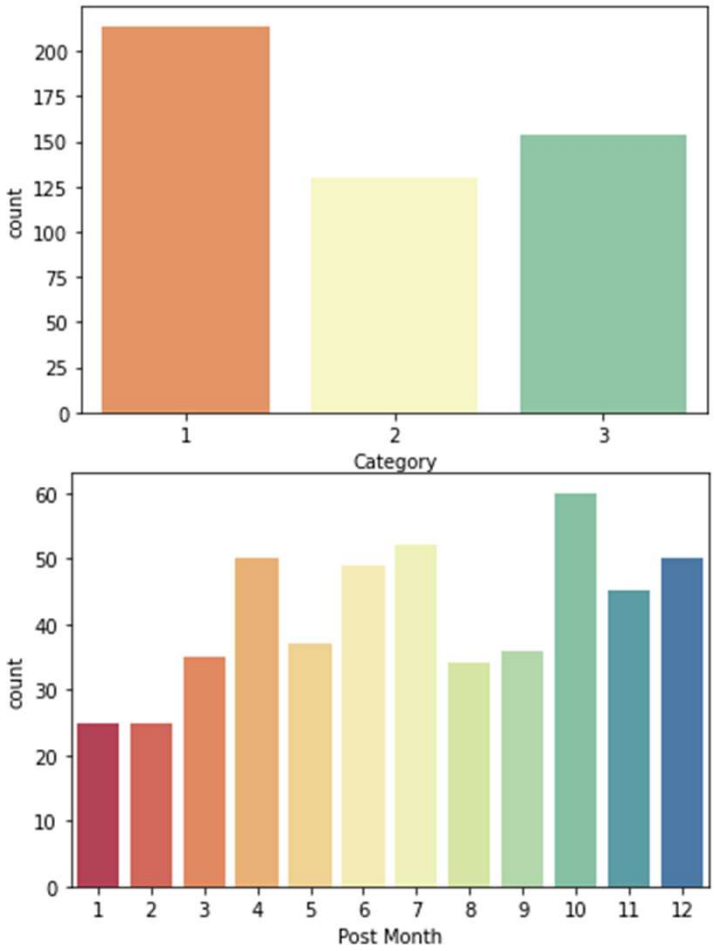
- (Moro et al., 2016) Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. Journal of Business Research, 69(9), 3341-3351.
- With codes taken from:

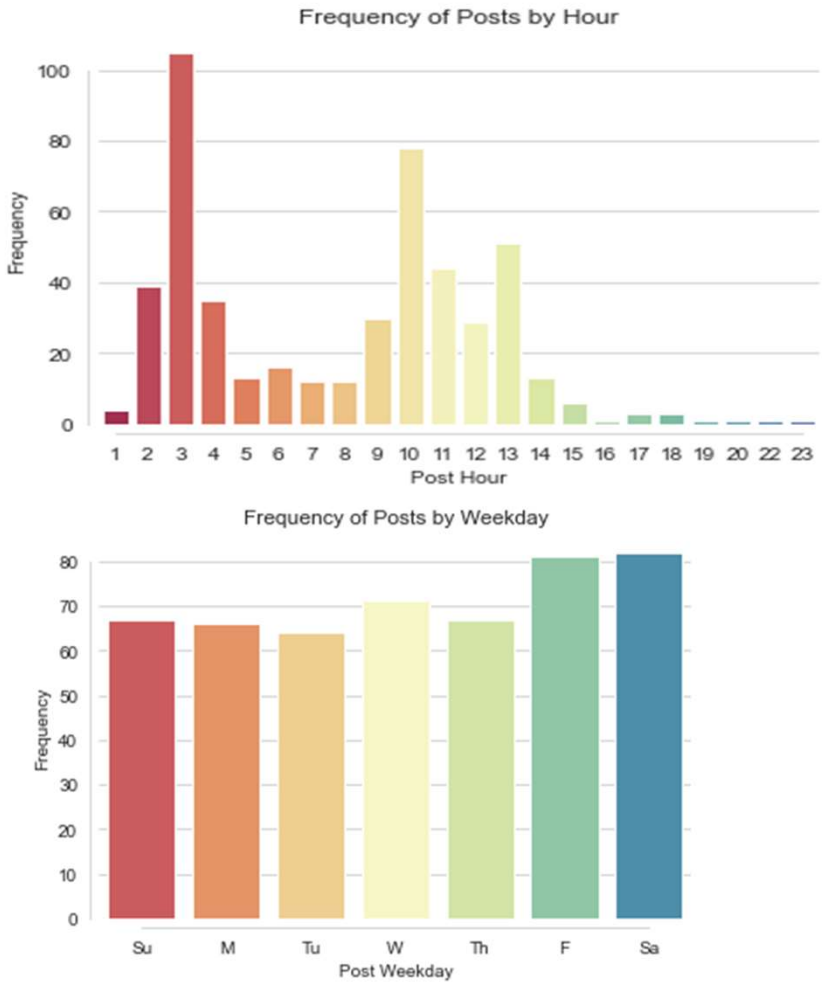
  https://github.com/jacobod/Facebook-Metrics-Prediction

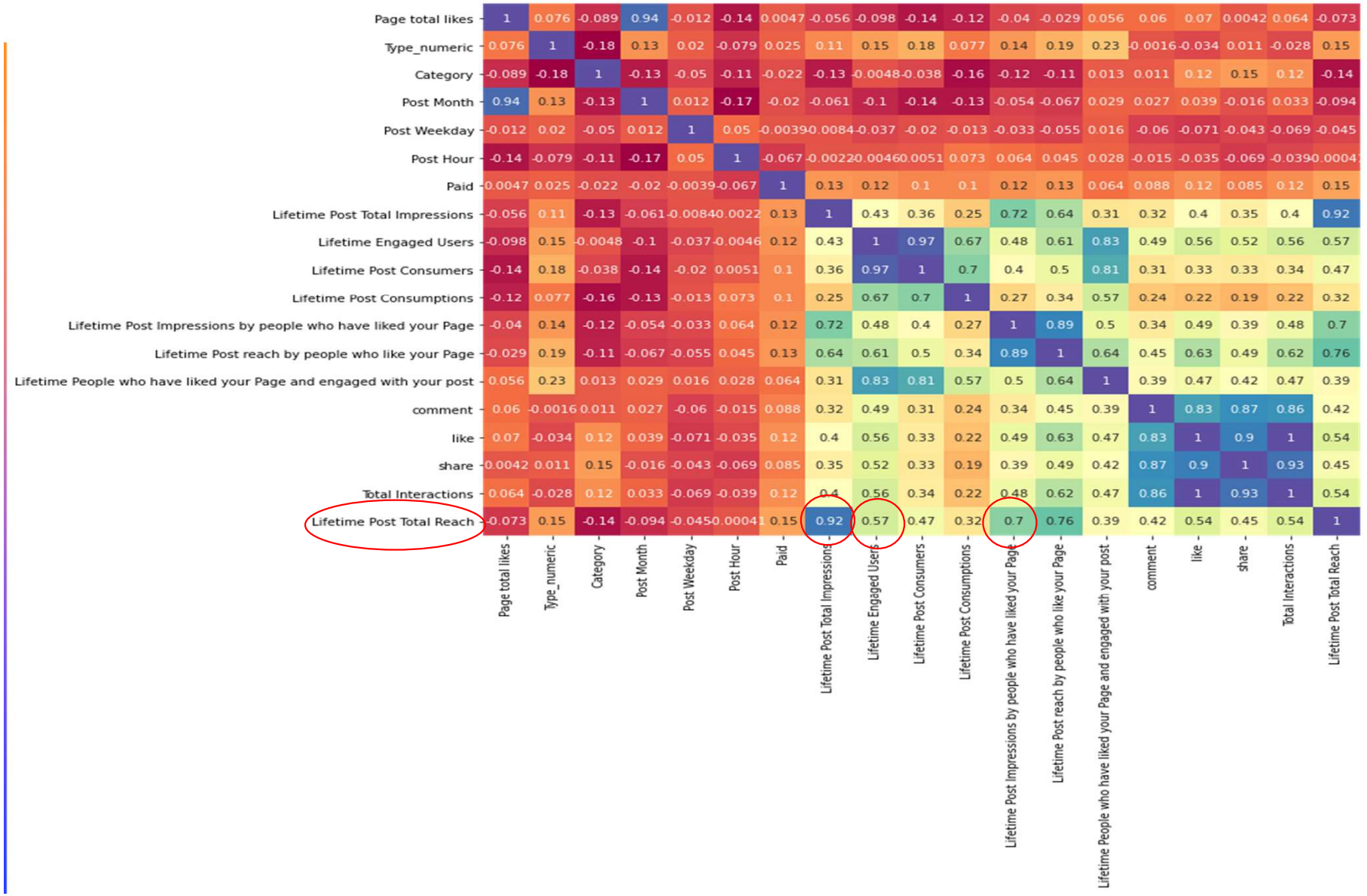
Process Workflow

1. EDA
    i. Analyse the frequency against various factors such as Category, Post Hour, Post Month, Post Weekday
2. Data preparation
    i. Removed outliers
    ii. Lifetime Post Total Reach as the y-variable
3. Data analysis
    i. Find out the correlation between factors
4. ML model training
    i. Simple Linear, Multiple Linear, and Random Forest Regression
5. Evalution

Some analysis






Page total likes 1 0.076 -0.089 0.94 -0.012 -0.14 0.0047 -0.056 -0.098 -0.14 -0.12 -0.04 -0.029 0.056 0.06 0.07 0.0042 0.064 -0.073
Type_numeric 0.076 1 -0.18 0.13 0.02 -0.079 0.025 0.11 0.15 0.18 0.077 0.14 0.19 0.23 -0.0016 -0.034 0.011 -0.028 0.15
Category -0.089 -0.18 1 -0.13 -0.05 -0.11 -0.022 -0.13 -0.0048 -0.038 -0.16 -0.12 -0.11 0.013 0.011 0.12 0.15 0.12 -0.14
Post Month 0.94 0.13 -0.13 1 0.012 -0.17 -0.02 -0.061 -0.1 -0.14 -0.13 -0.054 -0.067 0.029 0.027 0.039 -0.016 0.033 -0.094
Post Weekday -0.012 0.02 -0.05 0.012 1 0.05 -0.0039 0.0084 -0.037 -0.02 -0.013 -0.033 -0.055 0.016 -0.06 -0.071 -0.043 -0.069 -0.045
Post Hour -0.14 -0.079 -0.11 -0.17 0.05 1 -0.067 -0.0022 0.0046 0.0051 0.073 0.064 0.045 0.028 -0.015 -0.035 -0.069 -0.039 0.0004
Paid 0.0047 0.025 -0.022 -0.02 -0.0039 -0.067 1 0.13 0.12 0.1 0.1 0.12 0.13 0.064 0.088 0.12 0.085 0.12 0.15
Lifetime Post Total Impressions -0.056 0.11 -0.13 -0.061 -0.0084 0.0022 0.13 1 0.43 0.36 0.25 0.72 0.64 0.31 0.32 0.4 0.35 0.4 0.92
Lifetime Engaged Users -0.098 0.15 -0.0048 -0.1 -0.037 -0.0046 0.12 0.43 1 0.97 0.67 0.48 0.61 0.83 0.49 0.56 0.52 0.56 0.57
Lifetime Post Consumers -0.14 0.18 -0.038 -0.14 -0.02 0.0051 0.1 0.36 0.97 1 0.7 0.4 0.5 0.81 0.31 0.33 0.33 0.34 0.47
Lifetime Post Consumptions -0.12 0.077 -0.16 -0.13 -0.013 0.073 0.1 0.25 0.67 0.7 1 0.27 0.34 0.57 0.24 0.22 0.19 0.22 0.32
Lifetime Post Impressions by people who have liked your Page -0.04 0.14 -0.12 -0.054 -0.033 0.064 0.12 0.72 0.48 0.4 0.27 1 0.89 0.5 0.34 0.49 0.39 0.48 0.7
Lifetime Post reach by people who like your Page -0.029 0.19 -0.11 -0.067 -0.055 0.045 0.13 0.64 0.61 0.5 0.34 0.89 1 0.64 0.45 0.63 0.49 0.62 0.76
Lifetime People who have liked your Page and engaged with your post 0.056 0.23 0.013 0.029 0.016 0.028 0.064 0.31 0.83 0.81 0.57 0.5 0.64 1 0.39 0.47 0.42 0.47 0.39
comment 0.06 -0.0016 0.011 0.027 -0.06 -0.015 0.088 0.32 0.49 0.31 0.24 0.34 0.45 0.39 1 0.83 0.87 0.86 0.42
like 0.07 -0.034 0.12 0.039 -0.071 -0.035 0.12 0.4 0.56 0.33 0.22 0.49 0.63 0.47 0.83 1 0.9 1 0.54
share 0.0042 0.011 0.15 -0.016 -0.043 -0.069 0.085 0.35 0.52 0.33 0.19 0.39 0.49 0.42 0.87 0.9 1 0.93 0.45
Total Interactions 0.064 -0.028 0.12 0.033 -0.069 -0.039 0.12 0.4 0.56 0.34 0.22 0.48 0.62 0.47 0.86 1 0.93 1 0.54
Lifetime Post Total Reach -0.073 0.15 -0.14 -0.094 -0.045 0.00041 0.15 0.92 0.57 0.47 0.32 0.7 0.76 0.39 0.42 0.54 0.45 0.54 1
Page total likes
Type_numeric
Category
Post Month
Post Weekday
Post Hour
Paid
Lifetime Post Total Impressions
Lifetime Engaged Users
Lifetime Post Consumers
Lifetime Post Consumptions
Lifetime Post Impressions by people who have liked your Page
Lifetime Post reach by people who like your Page
Lifetime People who have liked your Page and engaged with your post
comment
like
share
Total Interactions
Lifetime Post Total Reach

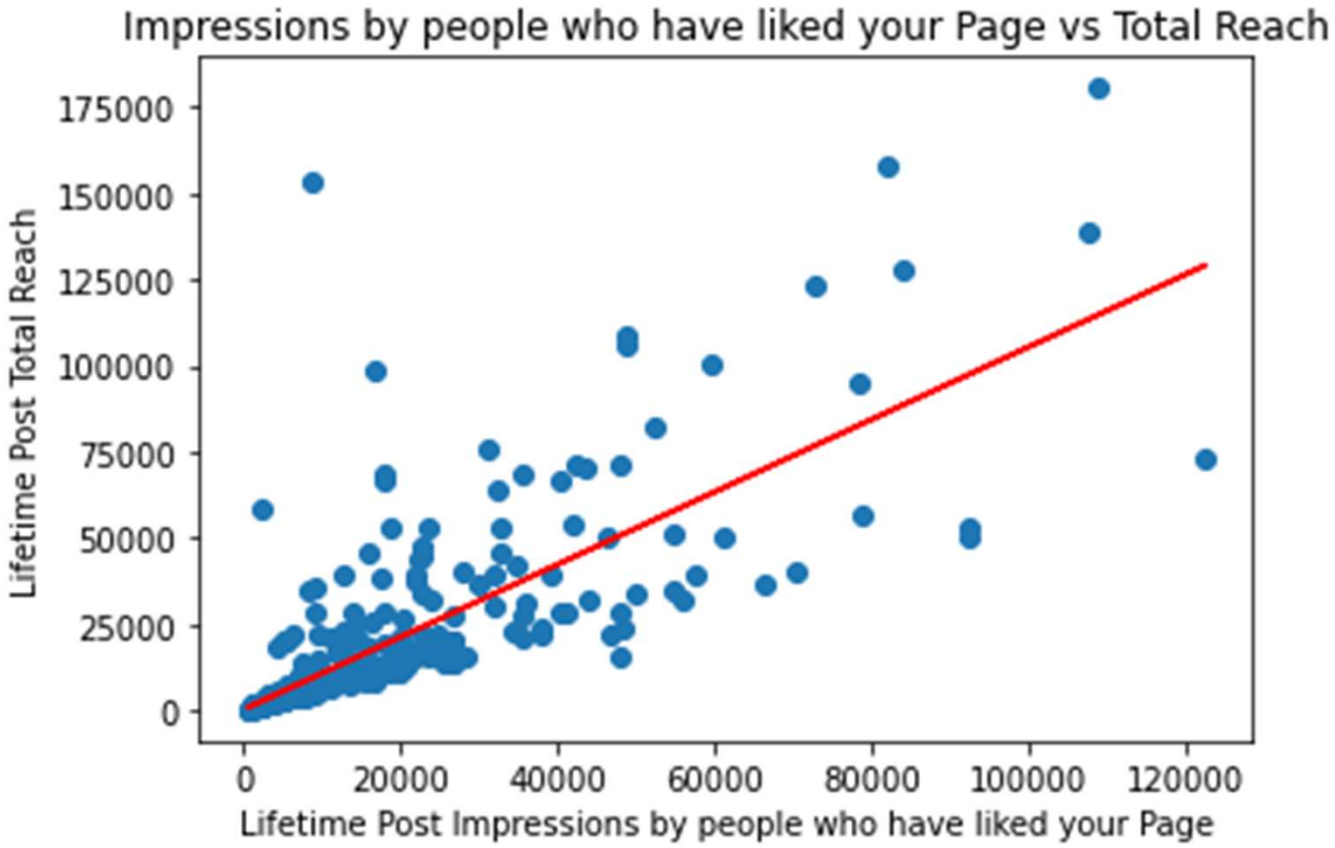
Machine Learning models

- Simple Linear regression
- Multiple Linear regression
- Random Forest regression

```
Regression model's training score = 0.62
Regression model's test score     = 0.68
```

test_size = 0.1

Simple Linear Regression



Multiple Linear and Random Forest Regression

- Lifetime Post Total Reach as the y-variable
- Lifetime Post Total Impressions, Lifetime Engaged Users, Lifetime Post Impressions by people who have liked your Page as the x-variables

Multiple Linear Regression Score

	Score	Step
0	0.892462	train
1	0.772500	test

test_size=0.25

Random Forest Regression Score

	Score	Step
0	0.944	train
1	0.822	test

test_size=0.4

Conclusions

- The team may consider factors with higher frequency to launch their promotions
    - Such as Category, Post Hour, Post Month, Post Weekday
- However, when measured alone, these factors have low correlation with Lifetime Post Total Reach
- The regression models explain the factors well
- Will including dummy variables improve the model?