CS 138 CAPSTONE A.Y 2023-2024

Travel Time Prediction Based on Metro Manila Traffic Density and Weather Conditions

Presented by **Newton-Cutes**

Background

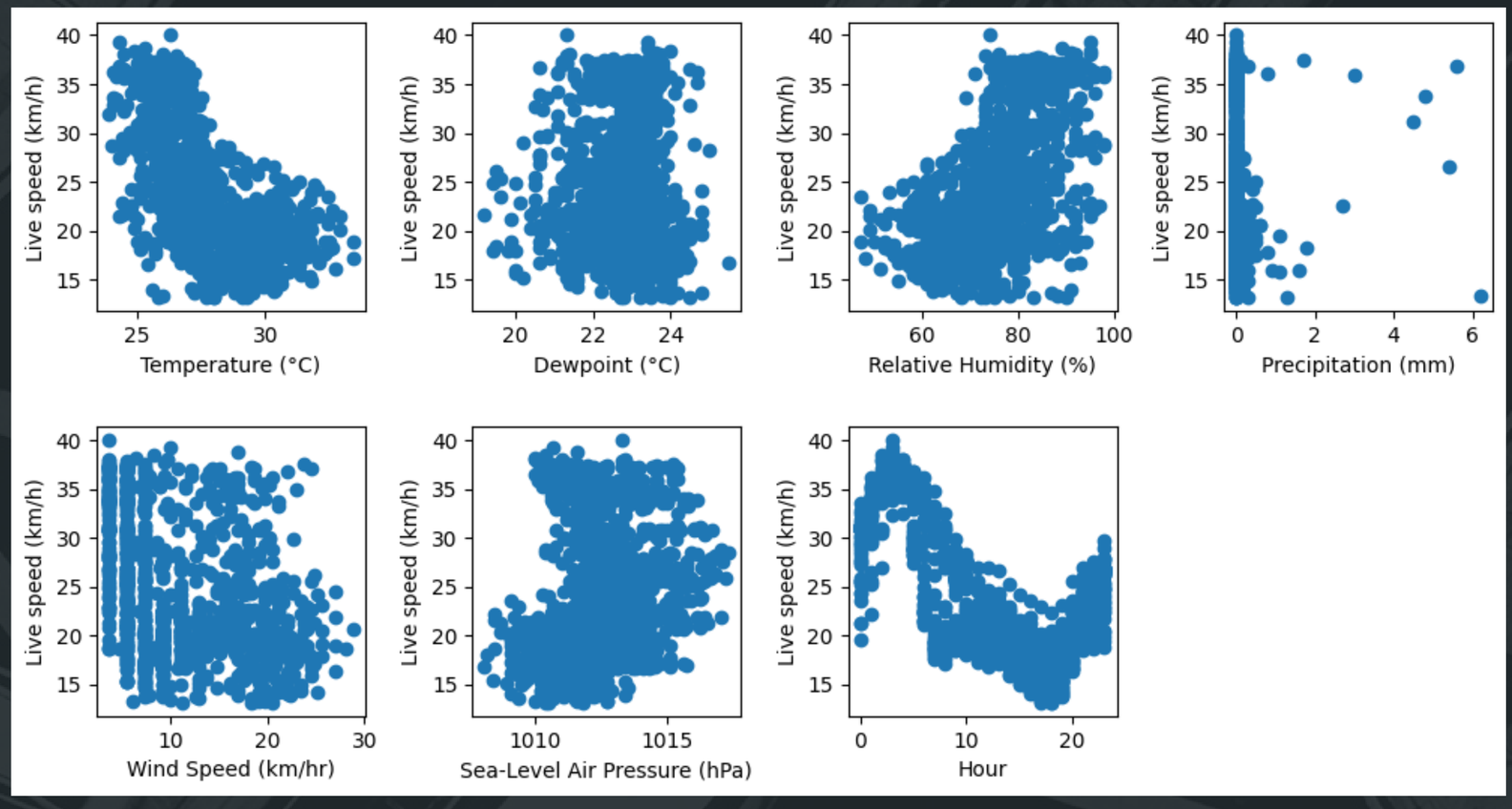
Metro Manila is one of the most congested places in the Philippines, and arriving to places on time presents a challenge to workers and students everyday. To better estimate their arrival times, some people use navigation apps, such as Google Maps and Waze.
To attempt to find a more accurate estimate, we consider one specific factor -- **weather**.

Preliminary Results

After collecting traffic data from Tomtoms and weather data from Meteostat [1] [2] from Nov. 26, 2023 to Jan. 13, 2024, we tested

each weather feature vs. live speed

The hour graph show that live speed peaks in the early hours and starts dropping midday. Live speed is inversely proportional to temperature.



Next, we trained and tested the dataset using polynomial regression models from degrees 1 to 4 to determine which method best fits the data [3].

1 degree model	
Training R2	0.6645214284148542
Testing R2:	0.5063973461422476
RMSE:	4.678842094821674
2 degree model	
Training R2	0.7631380079980539
Testing R2:	0.5577434961439001
RMSE:	4.428807049476733
3 degree model	
Training R2	0.850103898910991
Testing R2:	0.010075356169452765
RMSE:	6.625989045187333
4 degree model	
Training R2	0.9107182052238145
Testing R2:	-5.173709517307195
RMSE:	16.54713237516209

Table 1: Training and Testing Results

Results and Discussion

From the results of our testing and training, we found out that the

quadratic regression model showed the highest accuracy and lowest RMSE

among the four degrees of polynomial regression models used.

76.31% training & **55.77%** testing & **4.43** RMSE

Significance Analysis

Here are the results of the significance analysis on weather features by incrementing each attribute by percentages to see if it affects the outcome of the quadratic regression model:

- **time** has a **significant effect**
- **weather** has **no significant effect**

	R2 score of Linear Regression
Incl. temp, prcp etc.	0.7631380079980539
Only hour	0.6218086472708723

Table 2: Isolation of the Hour Attribute

Conclusion & Recommendations

- **Weather** - **no significant effect**
- **Weather + Time** - **not enough** to predict
- **Regression** is a **sufficient method** but **add more/other features**
- **include public vehicles** in the scope
- choose a **different location** (where weather conditions might be more significant)
- needs **more data**

Altered attribute	5%	10%	15%
temp	0.557743495931723	0.5577434960623371	0.5577434963404816
dwpt	0.5577434959640819	0.5577434962542067	0.5577434958809391
rhum	0.5577434957416614	0.5577434958115174	0.5577434962489942
prcp	0.5577434965880195	0.5577434963664685	0.5577434964023271
wspd	0.5577434964170989	0.5577434962563417	0.5577434962618588
pres	0.5577434959290986	0.5577434959808814	0.5577434963570325
Altered attribute	20%	25%	30%
temp	0.5577434960874156	0.5577434961161197	0.5577434960134293
dwpt	0.5577434961810059	0.5577434962580445	0.5577434963599753
rhum	0.5577434964871952	0.5577434962966143	0.5577434963114724
prcp	0.5577434963259958	0.5577434963613326	0.5577434965321622
wspd	0.5577434957083132	0.5577434964563166	0.5577434958528904
pres	0.5577434959497305	0.557743495572306	0.5577434963158777

Table 3: Significance Analysis

References:
[1] Tomtoms - tomtom.com [2] Meteostat - meteostat.net/en/station/98425 [3] Spanton. Polynomial Regression: An Introduction - builtin.com/machine-learning/polynomial-regression