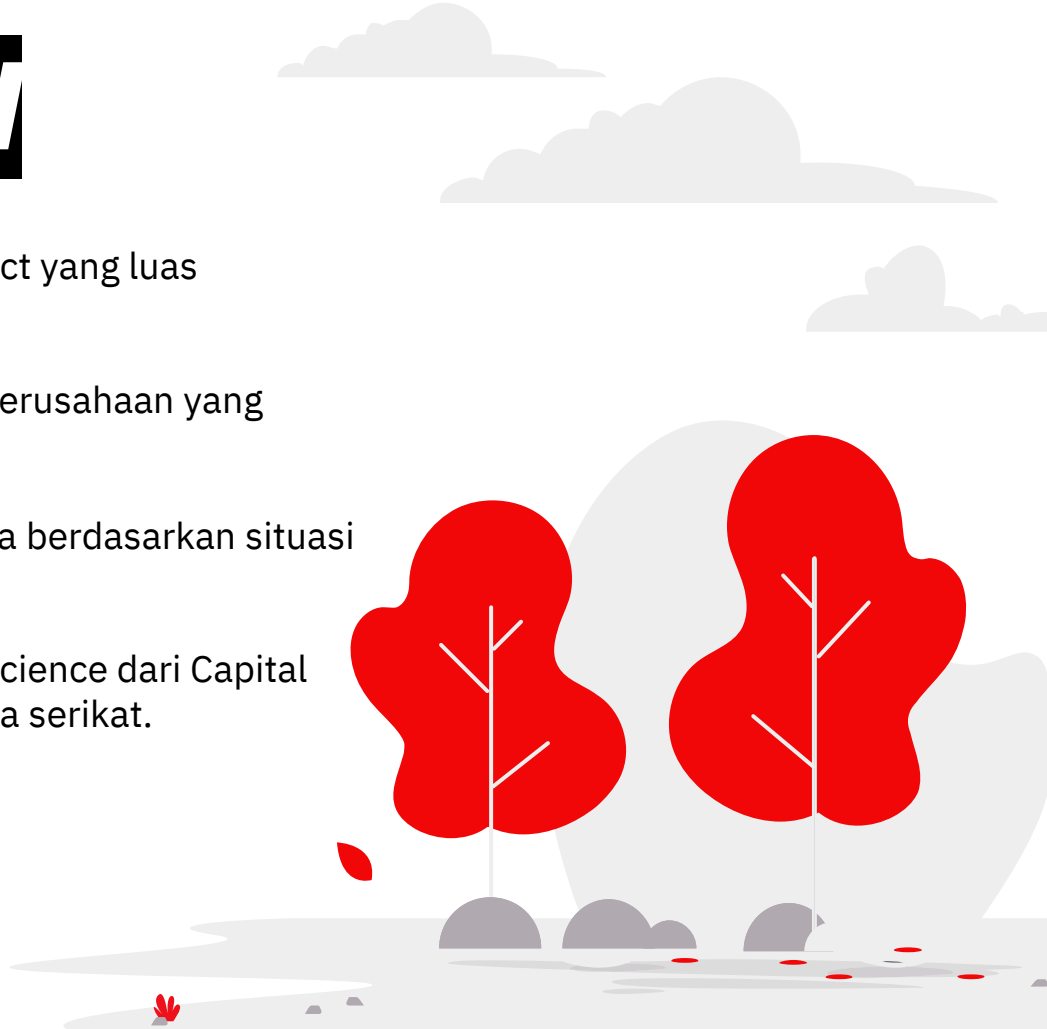
Epsilon Squad

BIKE SHARING DEMAND

Background

- Bikesharing merupakan bisnis dengan impact yang luas terhadap masyarakat.
- bisnis ini sangat menantang. Tidak sedikit perusahaan yang gagal dan tidak dapat bertahan lama.
- Analisis permintaan penyewaan dari sepeda berdasarkan situasi tertentu akan membantu perusahaan.
- Kami memposisikan diri sebagai Tim Data Science dari Capital Bike Share, sebuah perusahaan asal amerika serikat.



Problem Statement

Problem statement for Analytics :

- Seperti apa program yang sesuai dengan behavior dan segmen yang berbeda?

- Kapan dan dalam situasi apa program yang dijalankan dapat meningkatkan revenue?

Problem statement for Machine Learning :

- Bagaimana memprediksi banyaknya demand penyewaan sepeda sehingga kita bisa meminimalisir stock sepeda yang berlebih atau kehilangan konsumen karena stock sepeda tidak tersedia?

Goals

- **Memaksimalkan profit berdasarkan situasi yang dinamis.**
- **Menghindari potensi kehilangan pelanggan**
- **Mengurangi resiko bertambahnya biaya perawatan**

Data Understanding

- Data yang digunakan penyewaan Capital Bike Share pada tahun 2011 & 2012.
- Kami menambahkan juga data external.
- Data external yang dipakai disesuaikan dengan kondisi waktu agar relevan.

Data Understanding

Attributes Information

1. **dteday** : tanggal
2. **season**: Season / Musim
3. **yr** : Tahun
4. **Instant :** index
5. **mnth** : Bulan
6. **hr** : Jam
7. **holiday** : hari libur selain weekend(hasil ekstraksi Holiday Schedule Washington D.C)
8. **weekday** : Day of the week
9. **workingday** : Hari kerja
10. **casual** : Jumlah demand user non-member
11. **registered** : Jumlah demand user member
12. **cnt** : Total demand
13. **weathersit** : Kondisi cuaca
14. **temp** : Temperature (Celsius)
15. **atemp** : feels-like temperature (skala Celsius).
16. **hum** : Kelembaban.

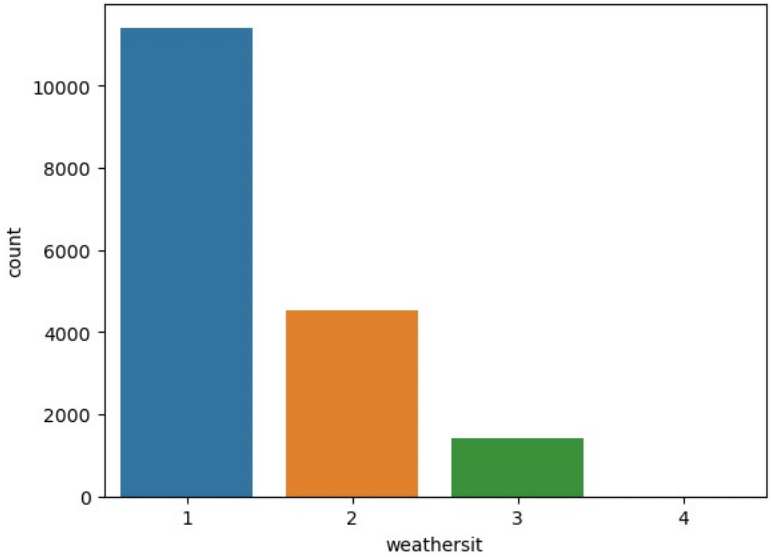
Data Understanding

Attributes Information (External data)

Data external yang ditambahkan disesuaikan, yaitu pada tahun 2011 dan 2012.

1. **Duration** : Durasi trip
2. **Start Date** : Keterangan waktu mulai trip
3. **End Date**: Keterangan waktu selesai trip
4. **Start Station**: Dock keberangkatan
5. **End Station** : Dock pengembalian
6. **Bike Number**: ID Sepeda
7. **Member Type** : Tipe konsumen Member/Non-Member

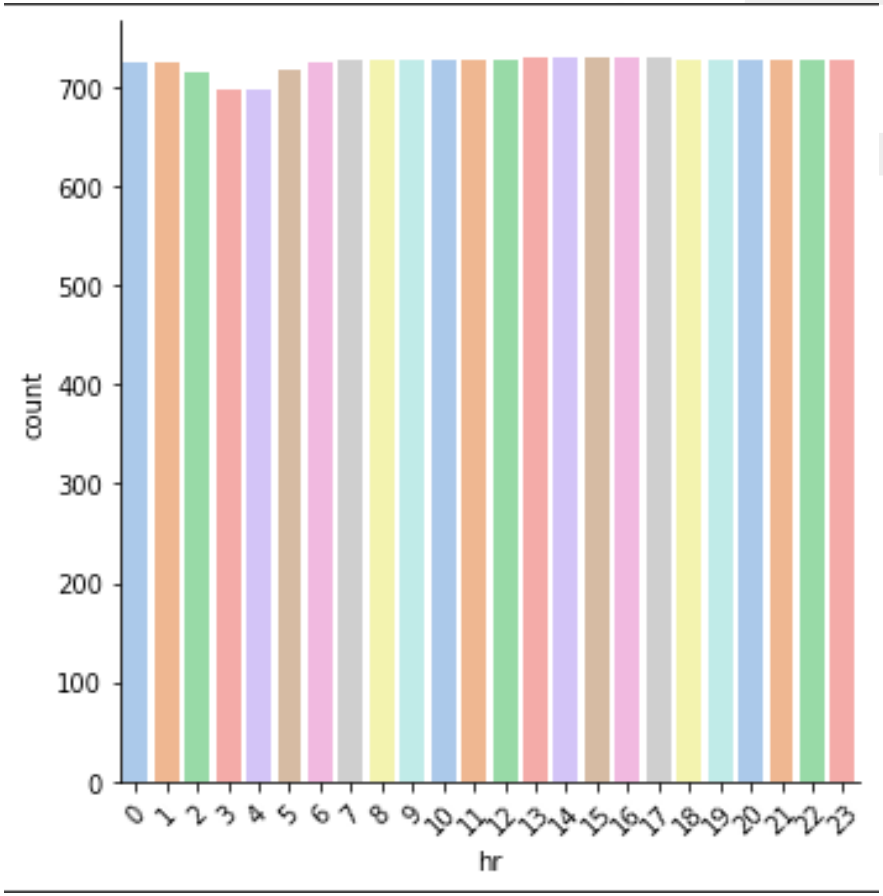
Exploratory Data Analysis



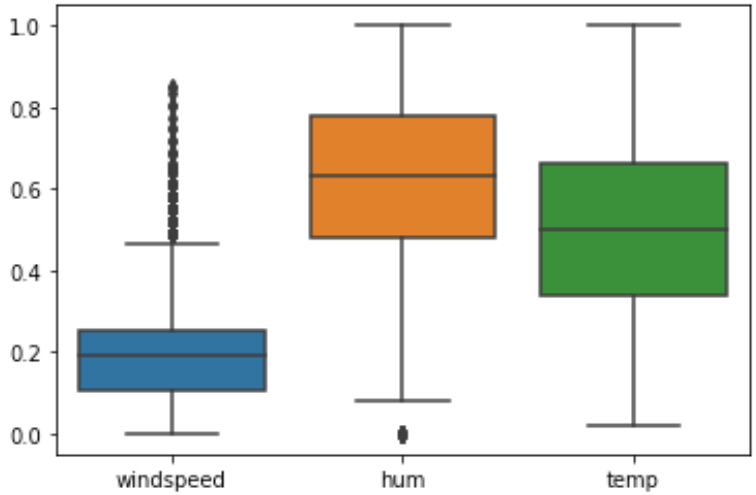
- Pada kolom **weathersit**; kondisi “Heavy rain” sangat kecil, hanya 3 data poin
- Kami masukan dalam kategori terdekat, yaitu “Hujan”

Exploratory Data Analysis
Missing Value

- Tidak ada Missing Value
- Hanya ada jam yang tidak tercantum



Exploratory Data Analysis
Outlier Detection

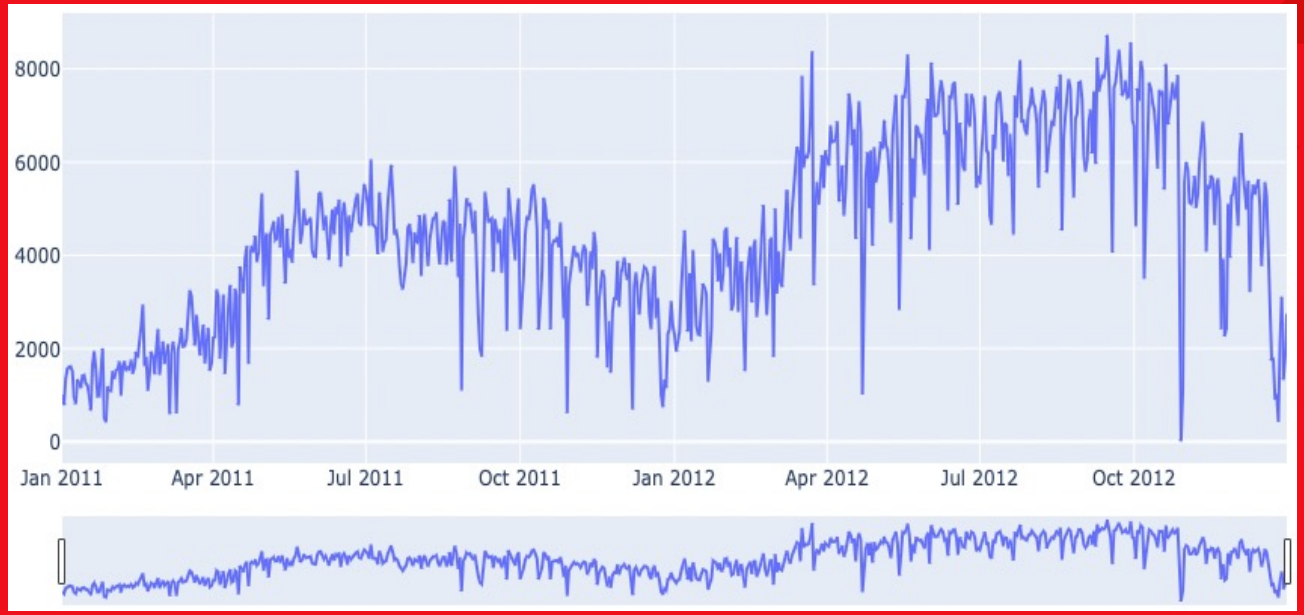


- Terlihat terdapat outlier di kolom windspeed dan hum

Data Analytics

Data Analytics

2011-2012 Trend



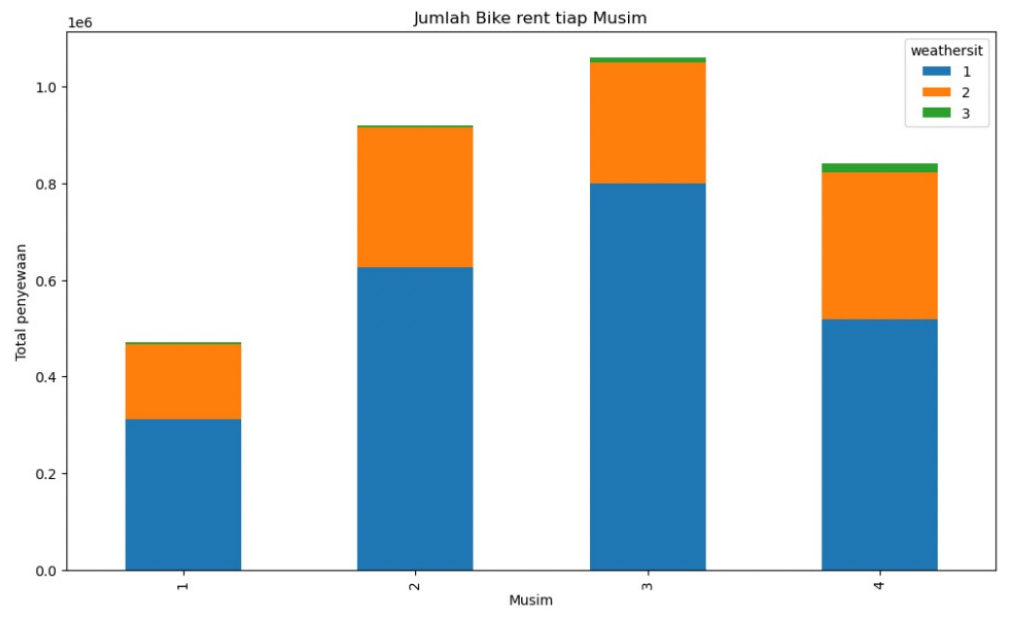
(2011-2012 Demand)

2011-2012

- Ada peningkatan demand dari 2011-2012
- Terlihat pergerakan yang berpola

Data Analytics

Demand sepeda tiap musim



- rental sepeda tertinggi pada Musim gugur & paling sedikit di musim semi.
- Penyewaan Ketika cuaca buruk sangat kecil.

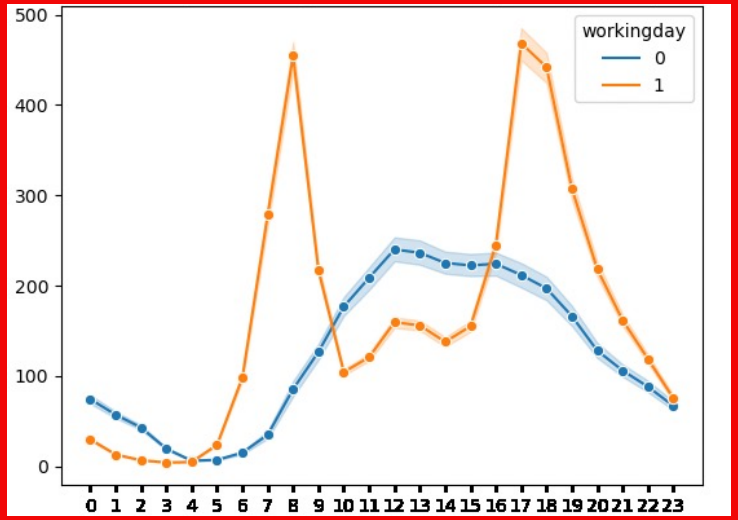
Weathersit :

- 1 = Cerah
- 2 = Berawan
- 3 = Hujan/Badai

Musim :

- 1 = Spring
- 2 = Summer
- 3 = Fall
- 4 = Winter

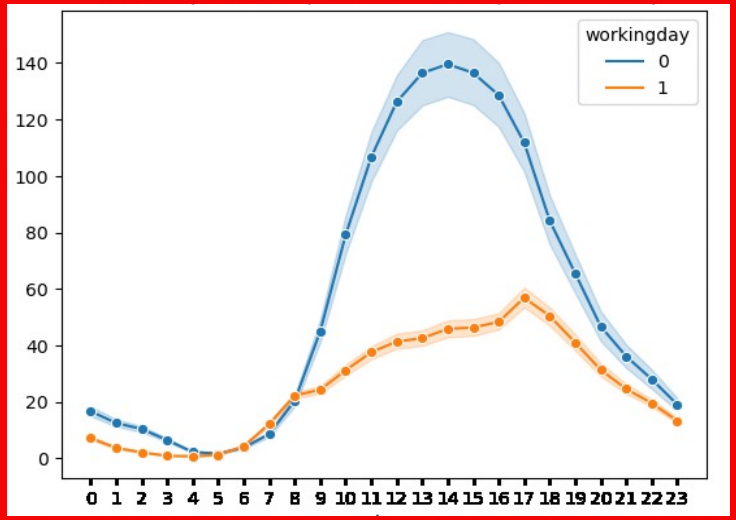
Data Analytics
Peak Hour Analysis



(Peak Hour Registered User)

- **registered consumer peak hournya ada pada jam 8 pagi dan jam 6 sore.**

Data Analytics
Peak Hour Analysis



(Peak Hour Casual User)

- **casual consumer ada pada siang hari.**

Data Analytics

Proporsi Demand

19%

casual user

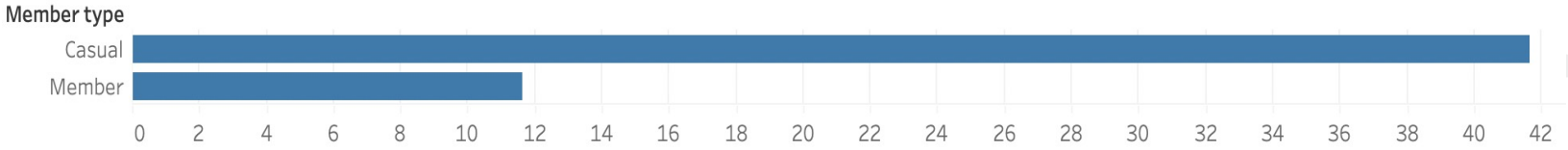
81%

registered member

(Persentase permintaan tiap tipe user)

Data Analytics

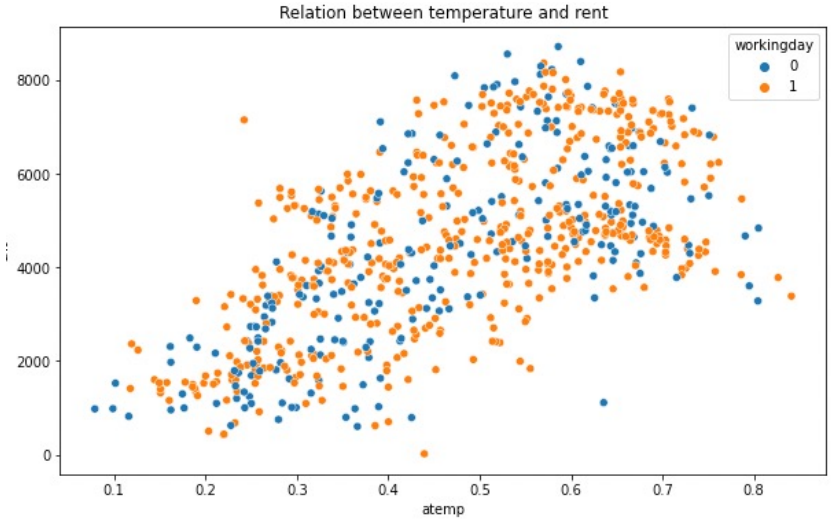
Durasi berkendara



- Durasi berkendara konsumer casual > registered user.
- registered -> sarana commuter. Casual user -> sarana olahraga/rekreasi.

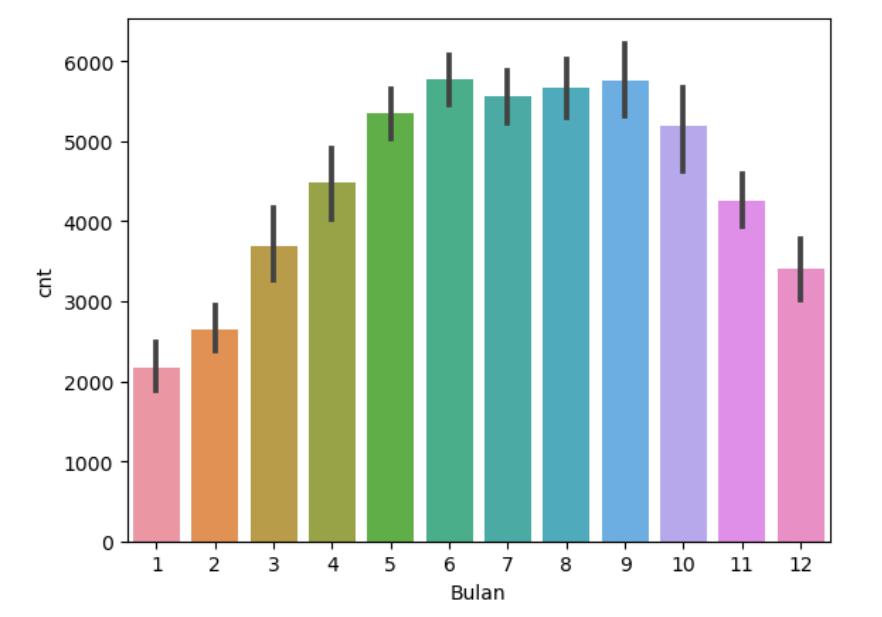
Data Analytics

Hubungan Suhu & Demand



- Suhu mempengaruhi banyaknya demand

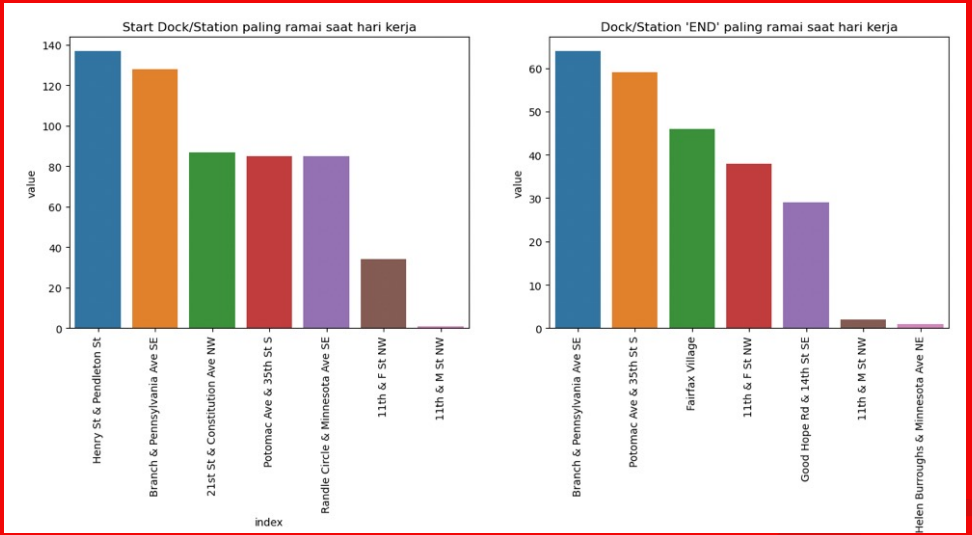
Data Analytics
Demand sepeda tiap bulan



- **Bulan-paling ramai terjadi pada pertengahan tahun.**
- **Terlihat penurunan ketika akhir dan awal tahun.**
- **Ada perbedaan signifikan saat bulan januari & Februari**

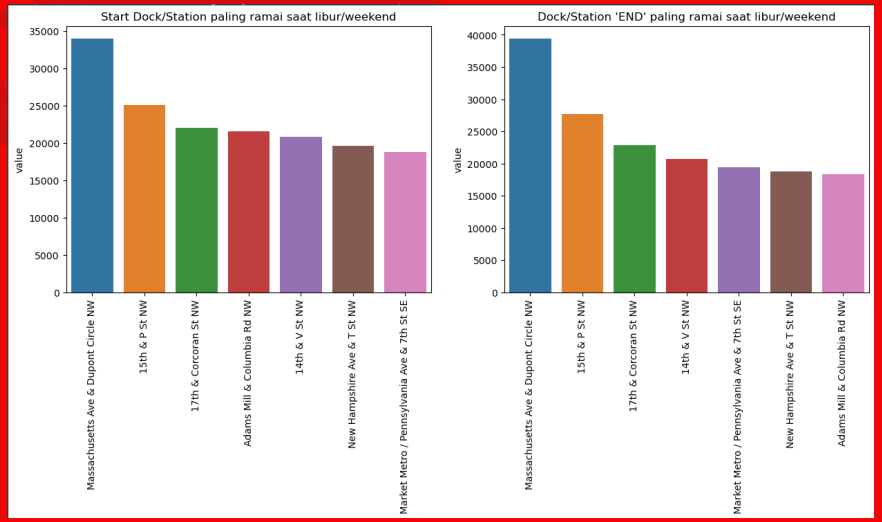
Data Analytics
Analysis Station/Docks

	value
count	193.000000
mean	5121.860104
std	6000.756751
min	1.000000
25%	729.000000
50%	2731.000000
75%	7781.000000
max	39416.000000



- Ada ketimpangan antara Docks yang terpasang.
- Satu dock bisa menerima 39.416 sepeda. Sedangkan ada yang hanya 1 kedatangan.
- evaluasi penempatan station/docks.

Data Analytics
Analysis Station/Docks



- Massachusetts Ave & Duport Circle menjadi stasiun tersibuk.
- Puncak kesibukan di Stasiun ada di hari libur.
- Pengembangan stasiun di daerah strategis akan berdampak positif.

Summary & Recommendations

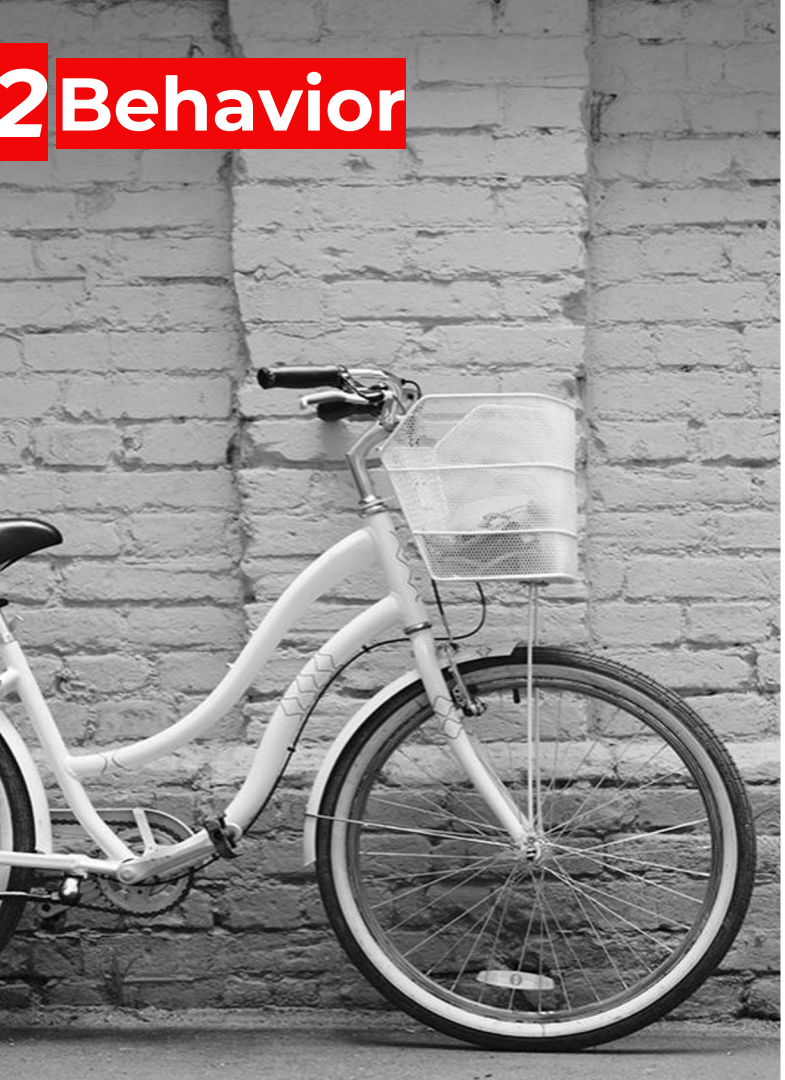
01 Efisiensi

Insight :

- Terdapat waktu tertentu demand berkurang signifikan.

Recommendation:

- Volume sepeda bisa dikurangi.
- Bulan-bulan sepi dapat dijadikan waktu maintenance tahunan.
- Kurangi stok sepeda di station sepi pada waktu tertentu.



2 Behavior

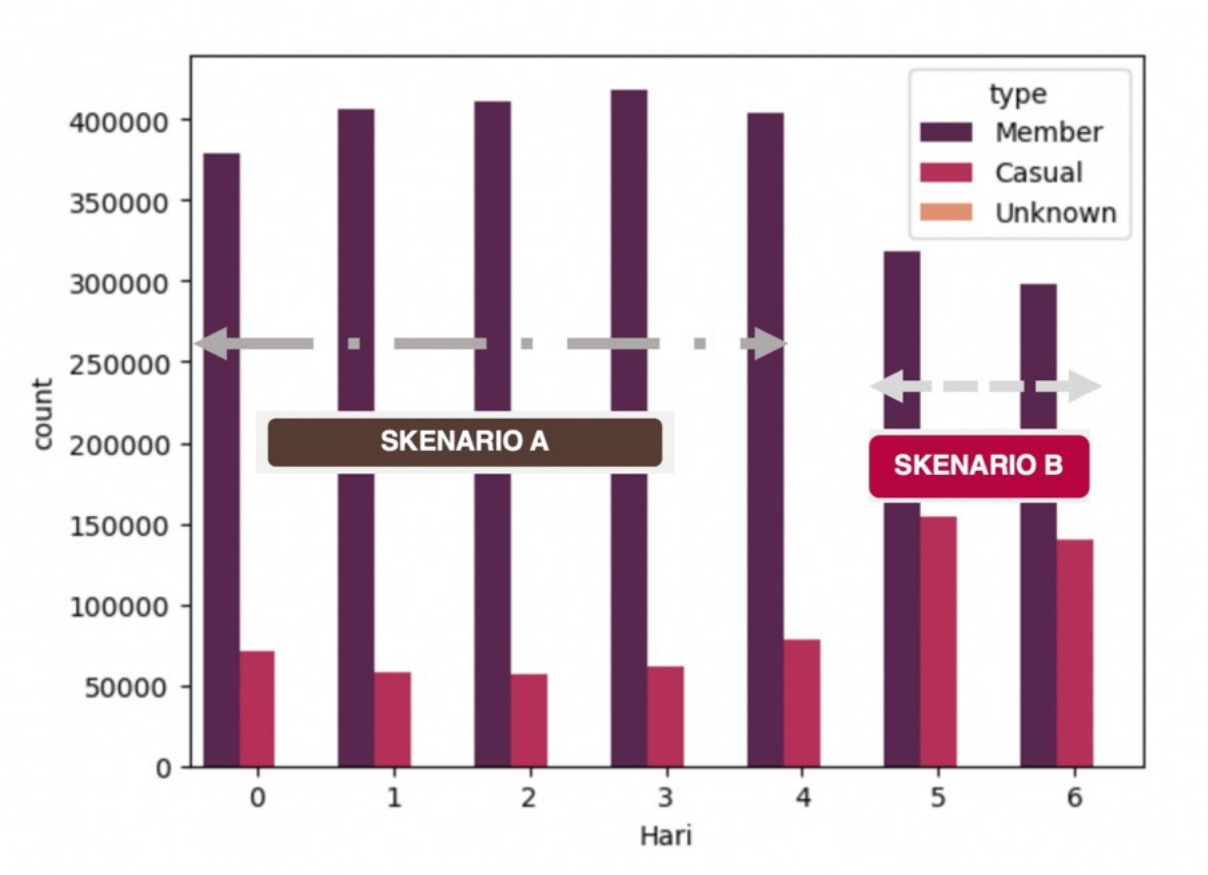
Insight :

- User terbagi menjadi dua segmen dengan behavior yang berbeda.

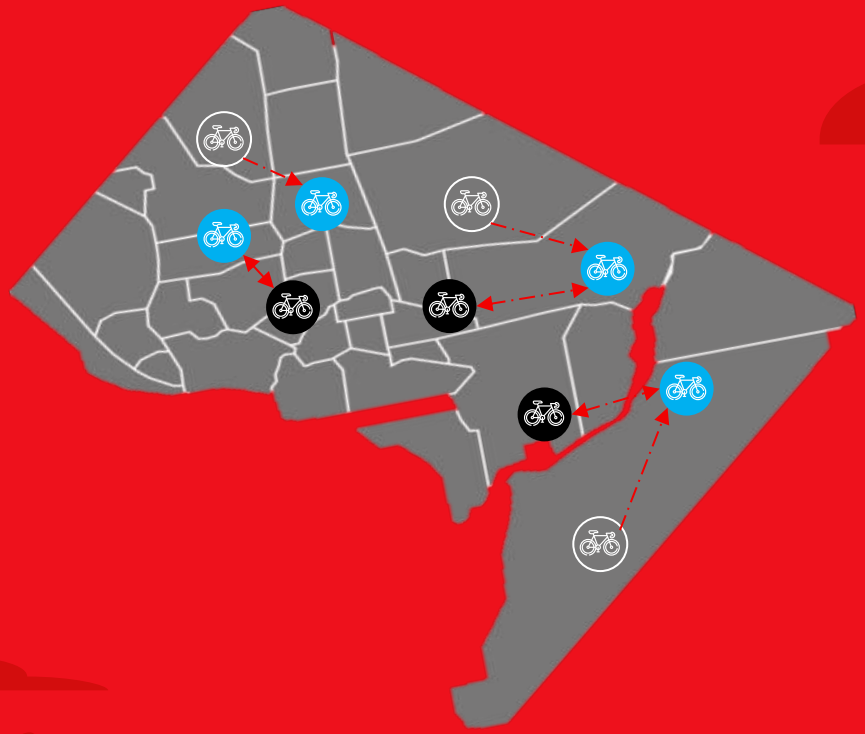
Recommendations:

- Pengembangan dock di titik strategis.
- kerjasama dengan dengan perusahaan disekitar rute dengan program bike to work.
- Pemberian promo pada waktu khusus seperti pada peak hour, liburan musim panas dan weekend

03 Skenario Re-Alokasi

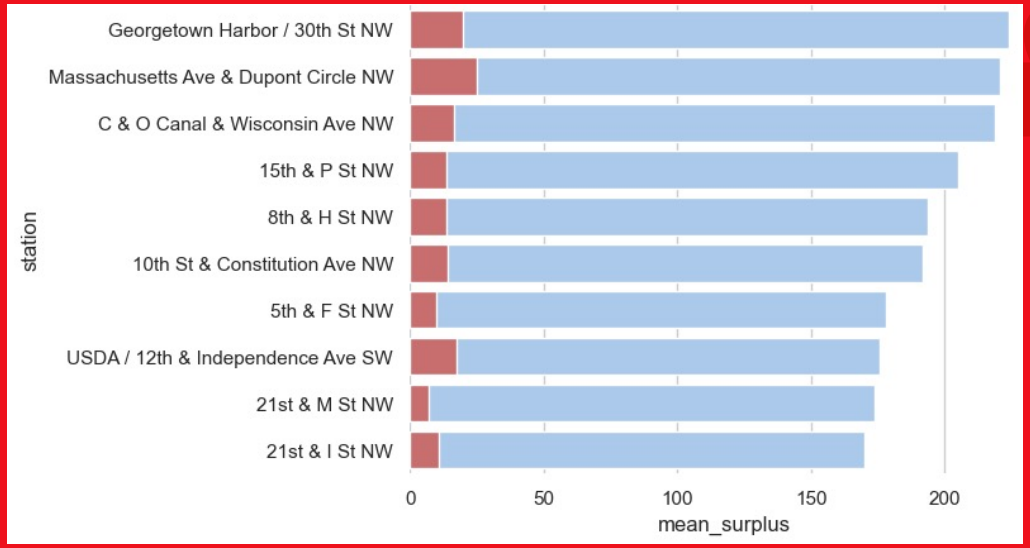


Skenario A



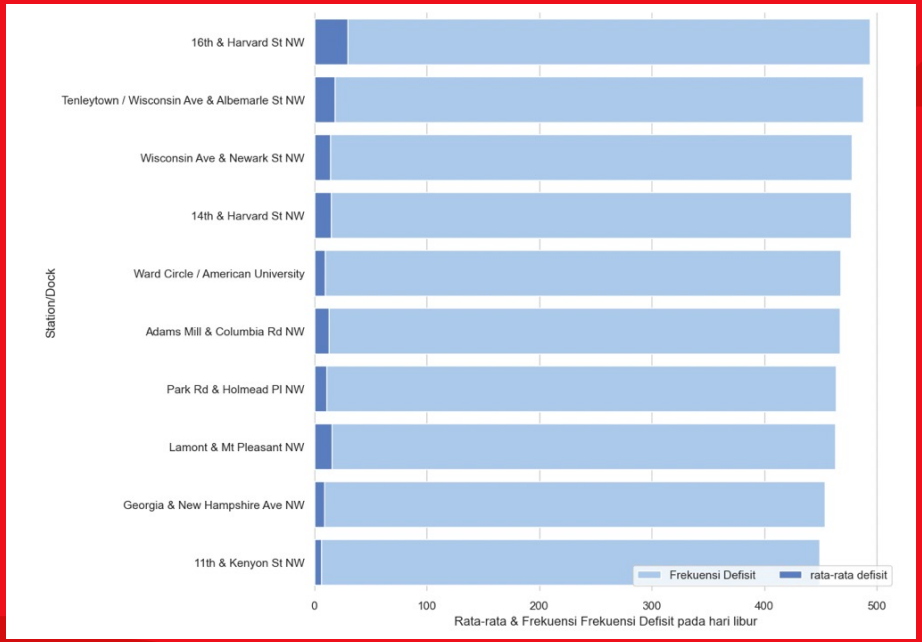
Skenario B

Skenario B



- Kebanyakan stasiun dengan stock sepeda berlebih berada di pusat kota

Skenario B



- Kebanyakan stasiun yang kekurangan sepeda berada di daerah permukiman

Machine Learning

Machine Learning

Feature Engineering

ENCODING

-One Hot Encoder
-Spline Transformer
-Polynomial Feature

Cross validation

TimeSeries Split

Scaling

Robust Scaler

Data Splitting

Train = 80%
Test = 20%

Machine Learning

Basic Model Building

Cross validation

TimeSeries Split

n_splits = 5
Gap = 48
Max_train_size = 1000
Test_size = 1000

Benchmark model

- Linear Regression
- Ridge
- Svr
- Enet
- Knn regressor
- Xgb regressor

Machine Learning

Basic Model Building

	Model	Mean_RMSE	Std_RMSE	Mean_MAPE	Std_MAPE
0	LinearR	-107.774721	13.983314	-0.600739	0.097023
1	Ridge	-108.237006	13.683174	-0.595972	0.097607
2	KNNR	-105.879671	21.601031	-0.504949	0.169661
3	XGBR	-93.069028	10.143670	-0.342203	0.049263
4	svr	-86.666137	9.111367	-0.528166	0.137176
5	enet	-243.460666	53.474712	-2.977934	0.893733

- XGboost dan SVR menjadi 2 kandidat teratas.
- Selanjutnya, kami bandingkan dengan memprediksi pada Test Set

Machine Learning

Predict to Test Set

	RMSE	MAPE
XGB	98.072714	0.617815
svr	217.782254	2.750528

- Nilai dari XG Boost Regressor lebih baik.
- model tersebut yang akan dilakukan hyperparameter tuning.

Machine Learning

Model Tuning

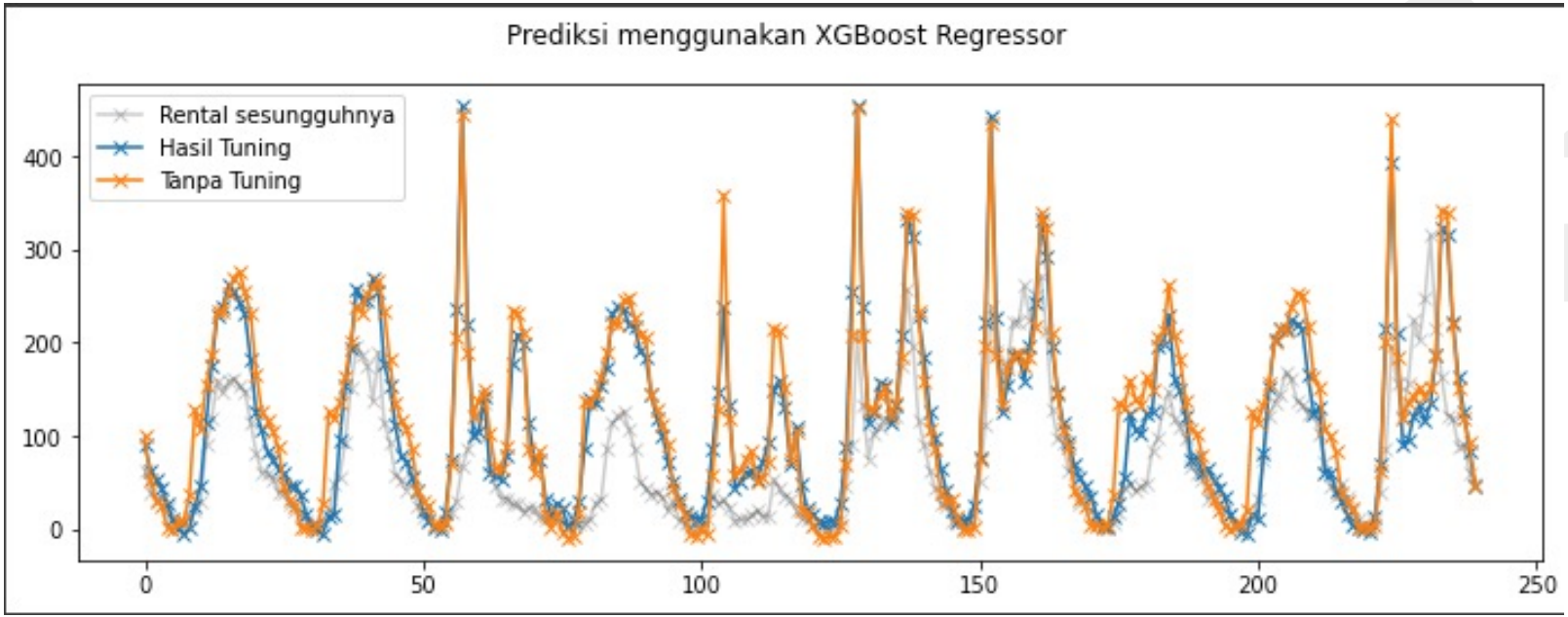
Hyperparameter :

- Max_depth
- Learning rate
- n_estimators
- Subsample
- Gamma
- Colsample_bytree
- Reg_alpha
- Min_child_weight

RandomizedSearchCV :

- N_iter = 50
- Random_state = 2022

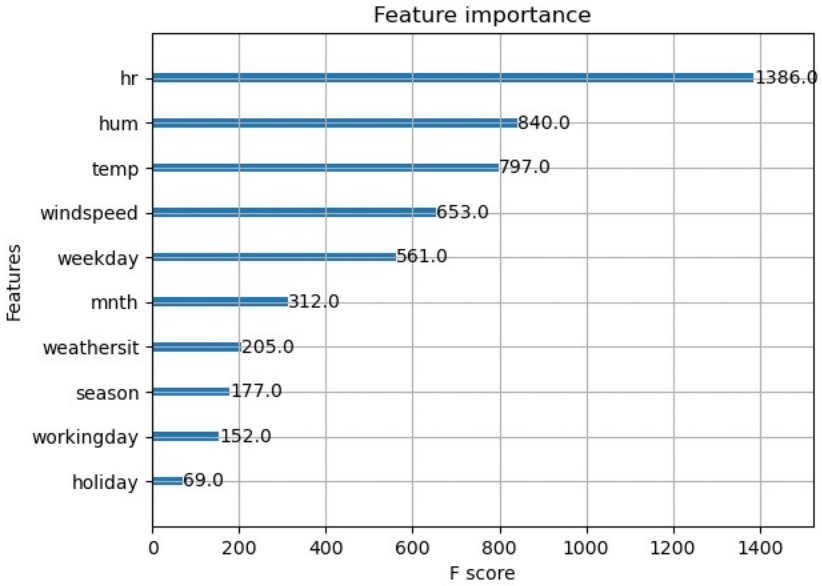
Before & After Tuning



	RMSE	MAPE
Before	79	0.50
After	73	0.46

Machine Learning

Feature Importances



- Hour, temp, humidity menjadi feature dengan kontribusi tertinggi dalam perubahan jumlah demand sepeda

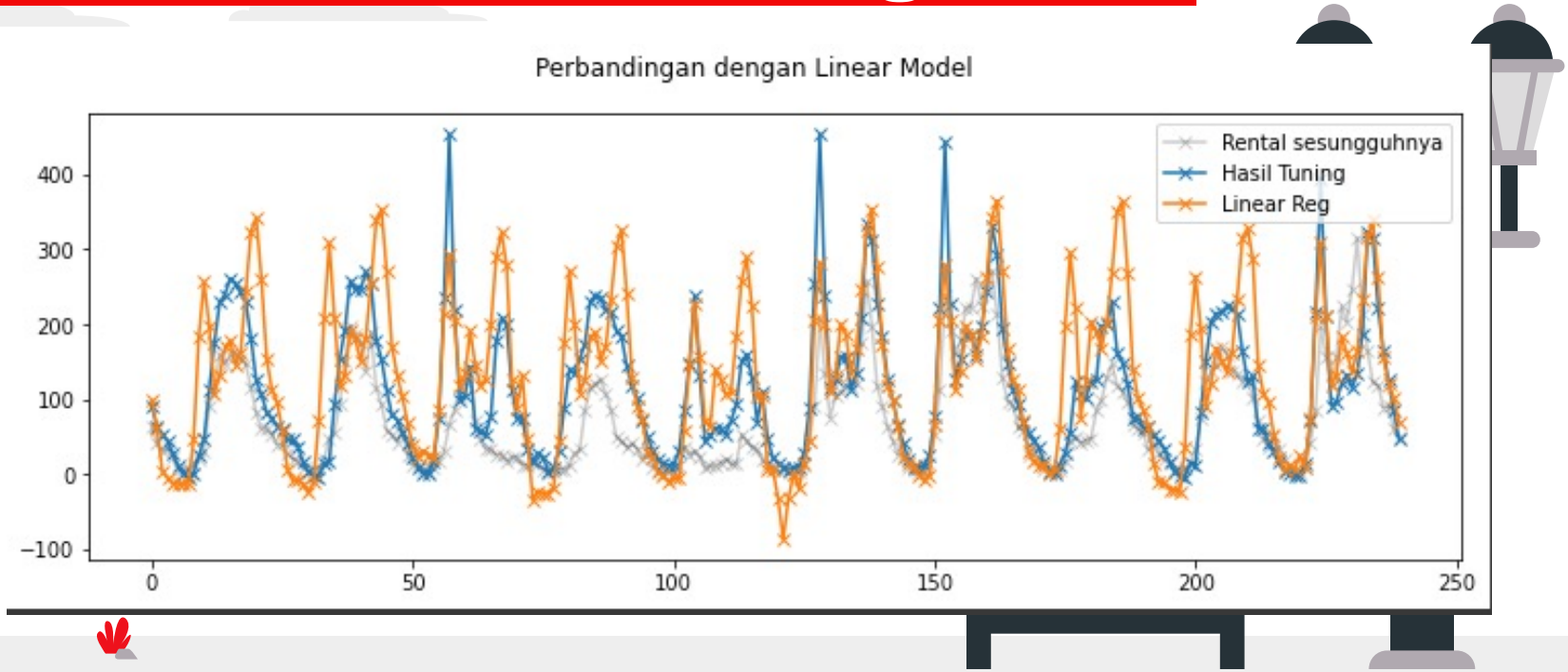
Evaluation Metric

- Metrik evaluasi yang adalah RMSE & MAPE.
- Hasil RMSE : 79

Conclusion

- asumsikan perusahaan sebelumnya menerapkan model linear regression (RMSE = 112).
- Perusahaan akan menerima benefit dengan selisih 33 unit.

Perbandingan model kami dengan actual demand & linear regression



Conclusion

- median durasi = 10 menit.
- harga dasar 0,05 dollar/menit.
- Maka opportunity cost setiap sepedanya adalah 3 dollar/perjam.
- Selisih dari model kami adalah 33
- 33 X $3 = 99 dollar.

Perusahaan bisa menghindari kerugian sebesar 99 dolar/jamnya. Maka dengan mempertimbangkan jam sibuk 10 jam/hari, perusahaan akan menghindari kerugian sebesar :

$990/hari

Limitasi model

- Bila demand sepeda perjamnya diluar rentang, model tidak dapat dipercaya
- Data yang kami gunakan observasi tahun 2011 – 2012, tidak relevan bila digunakan masa sekarang
- Masih terdapat outlier dalam model

Pengembangan model

- Menggunakan data terbaru yang dimiliki Capital Bike Share
- Total biaya penyewaan
- Menambahkan fitur yang lebih detail :
  - Durasi
  - Jarak tempuh
  - Jarak dock dengan moda transportasi lain
  - Latitude & Longitude

Terima Kasih .