BIKE SHARING DEMAND PREDICTION

By :
Syahlan Ramadhan
DTI-DS 0106

BIKE SHARING IN THE WORLD

10 million shared bikes and 3,000 bike sharing systems company exist across the world in 2021.

Chinese bike share firm goes bust after losing 90% of bikes

What we do in US ?

Data from : PBSC Urban Solutions & BCC

THE OVERVIEW

PROBLEM UNDERSTANDING

1. Which variables are significant in predicting the demand for shared bikes ?

2. How well those variables describe the bike demands ?

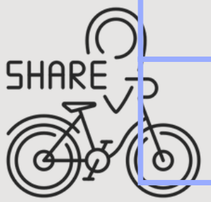
3. How the variable can impact to revenue and profit ?

GOALS

Know the best variable model can used to predict the demand for shared bake. and that variable choosen describe well when apply it. it is also can give

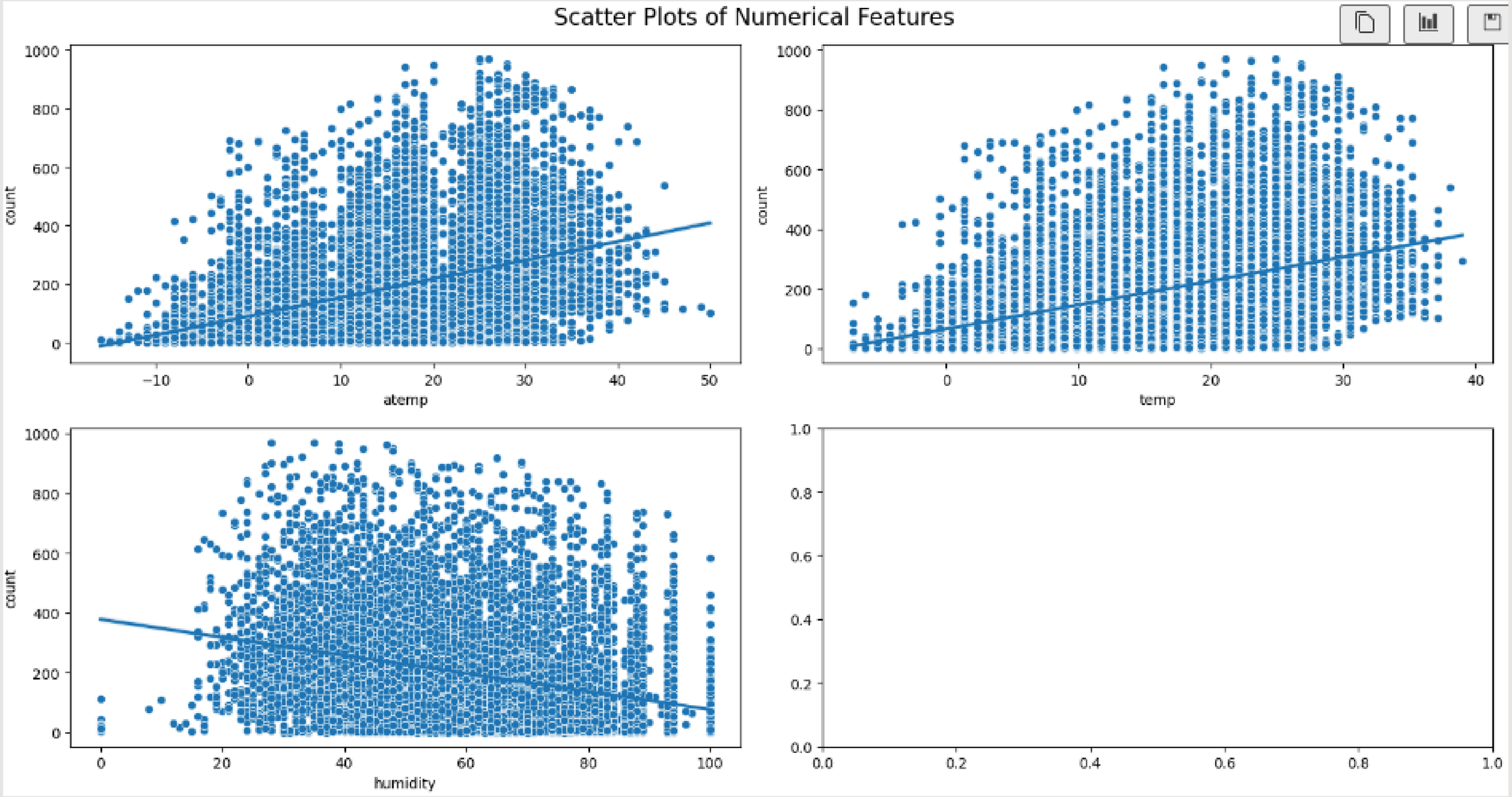
DATA UNDERSTANDING

Feature	Explanation	Null Count
dteday	date	non-null
season	1: winter, 2: spring, 3: summer, 4: fall	non-null
holiday	holiday or not	non-null
temp	normalized temperature in Celsius.	non-null
atemp	Normalized feeling temperature in Celsius	non-null
hum	normalized humidity	non-null
casual	count of casual bike users	non-null
registered	count of registered bike users	non-null
weathersit	1: Clear, 2:Mist, 3: Light snow, 4: Heavy rain	non-null

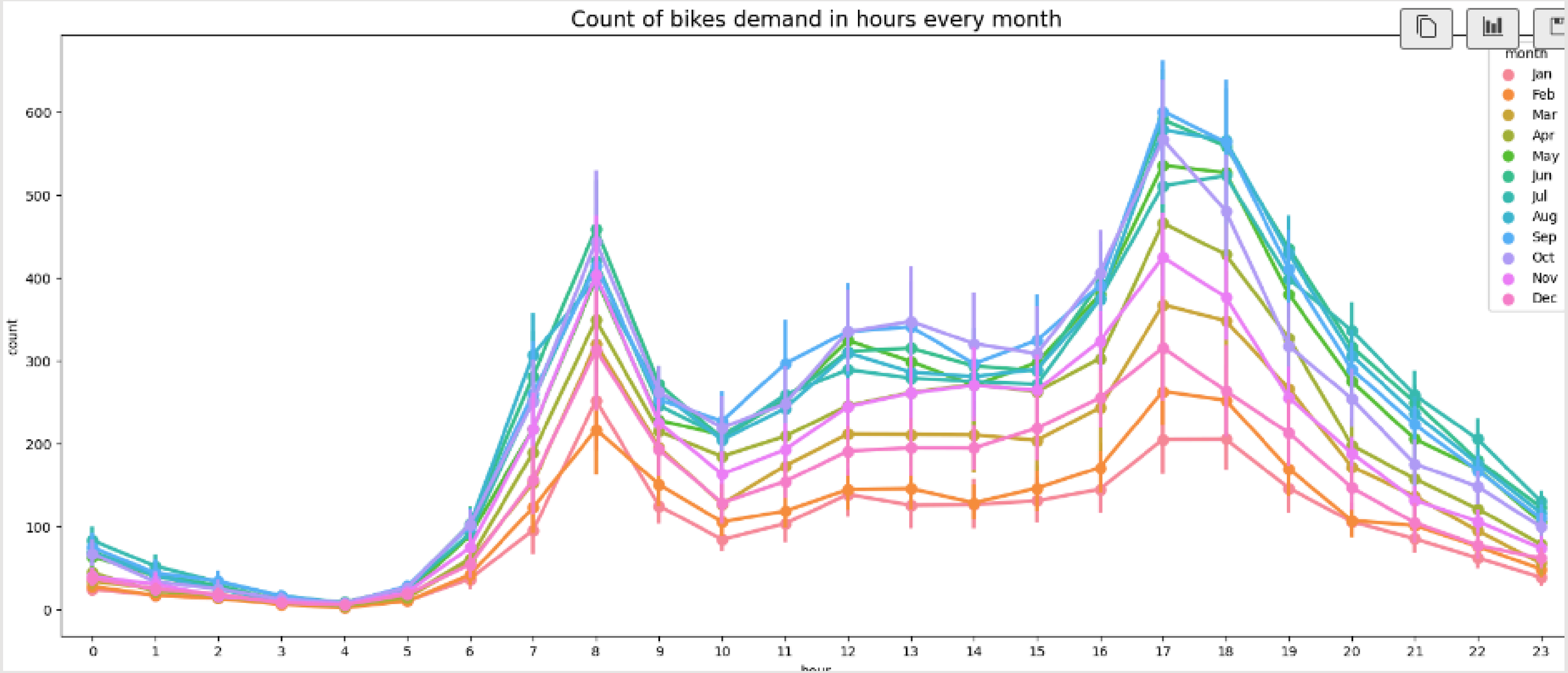


Target --> count (count of total rental bikes including both casual and registered)

DATA ANALYSIS



DATA ANALYSIS BIKE DEMAND IN HOURS EVERY MONTH



METRIC VALUATION

MAPE

is a metric that measures the accuracy of a predictive model. It calculates the average absolute percentage difference between the predicted and actual values, giving an indication of how far off predictions are in percentage terms.

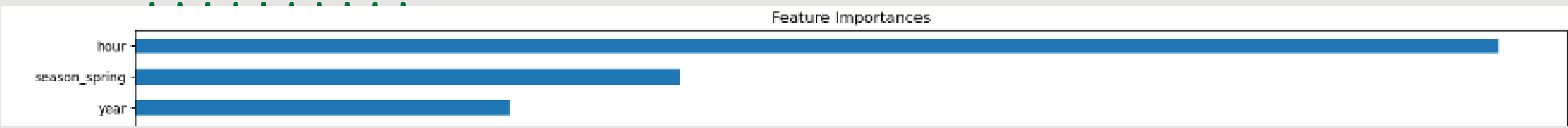
R- SQUARE

is a statistical metric that indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where a higher value indicates a better fit of the model to the data.

MODELING

Model	test mape	test r2	train mape	train r2	diff mape	diff r2
XGBRegressor	0.243256	0.948738	0.254014	0.938314	0.010758	0.010424
Random Forest	0.262272	0.946366	0.282810	0.928309	0.020538	0.018057
ADA Boost	0.274973	0.928541	0.291648	0.923261	0.016675	0.005280
Decision Tree	0.389213	0.883687	0.400117	0.866649	0.010904	0.017038
GB regressor	0.391359	0.809871	0.395985	0.802523	0.004626	0.007348
linear Regression	1.394.327	0.258882	1.385.671	0.213631	0.008655	0.045252

FINAL MODEL



Model Selection

XGBOOST REGRESSOR (AS HIGHEST SCORE MODEL)

After tuning:

- *mape score : 0.2313*
- *r2 score : 0.9583*

GB REGRESSOR (AS STABLE MODEL)

After tuning :

- *mape score : 02331*
- *r2 score : 0.9542*

Final Model

XGBOOST REGRESSOR (AS HIGHEST SCORE MODEL)

After feature selection :

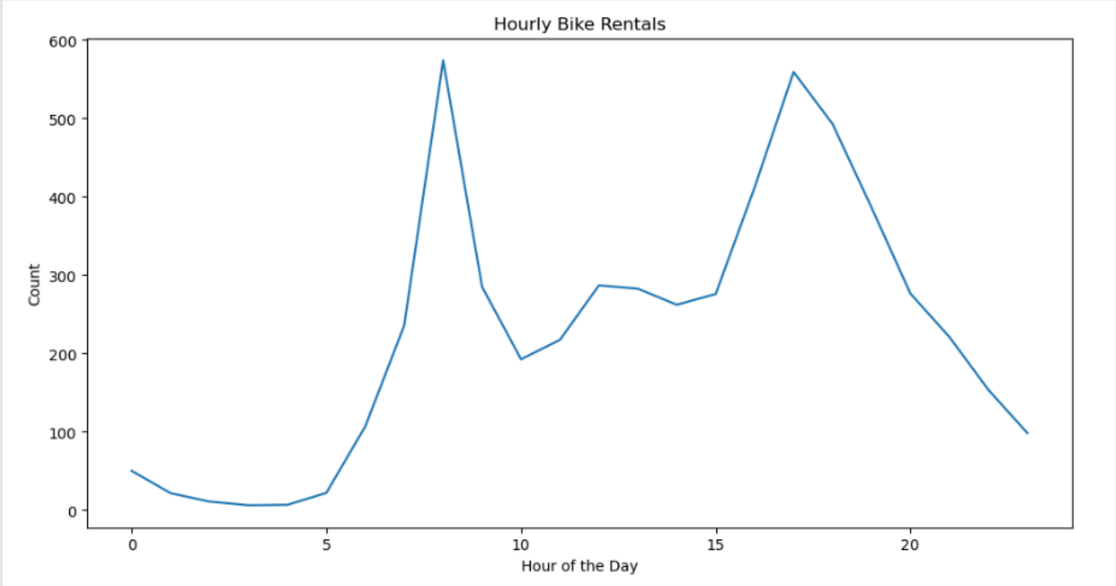
- *mape score : 0.2280*
- *r2 score : 0.9565*

CONCLUSION

3 VARIABLE IMPACT TO BIKE DEMAND

HOUR, SEASON, AND YEAR

ASSUME THAT :
500 BIKES PROVIDE BY COMPANY



NEXT YEAR PREDICTION

INCREASE ALMOST 100 BIKES
DEMAND PREDICTION IN MORNING HOUR:
05.00-08.00

ASSUME THAT :
$ 5 REVENUE /HOUR
$2 OPS COST /HOUR

TOTAL REVENUE: $ 2.317.884
TOTAL OPERATIONAL COST: $ 927.153

NET PROFIT: $ 1.390.730
EVERY QUARTAL

READ MORE ON MY GITHUB : HTTPS://GITHUB.COM/SYAHLAN-R/BIKE-SHARING-DEMAND

RECOMMENDATION

In-depth analysis to understand the causes of large underestimates or overestimates. Review whether there are external factors not included in the model, changes in user behavior, or policy changes that could affect demand patterns.

Need to prepared more bike in the morning to risk mitigation from lose of the customer. it’s can evaluate partialy every week by customer demand in the reality

Make a allocated cost timeline(ex: for repairing baik. employement, as well as maintenance in every point in company) to give the cost effecient to the company. so the company did not wasting their money for maintaining the bike

AND CONTINUE WITH DEPLOYMENT IN CLOUD