LinkedIn

slava.zagriichuk@gmail.com

COURIERS ONLINE PREDICTION

WITH AI

WHAT DO WE HAVE?

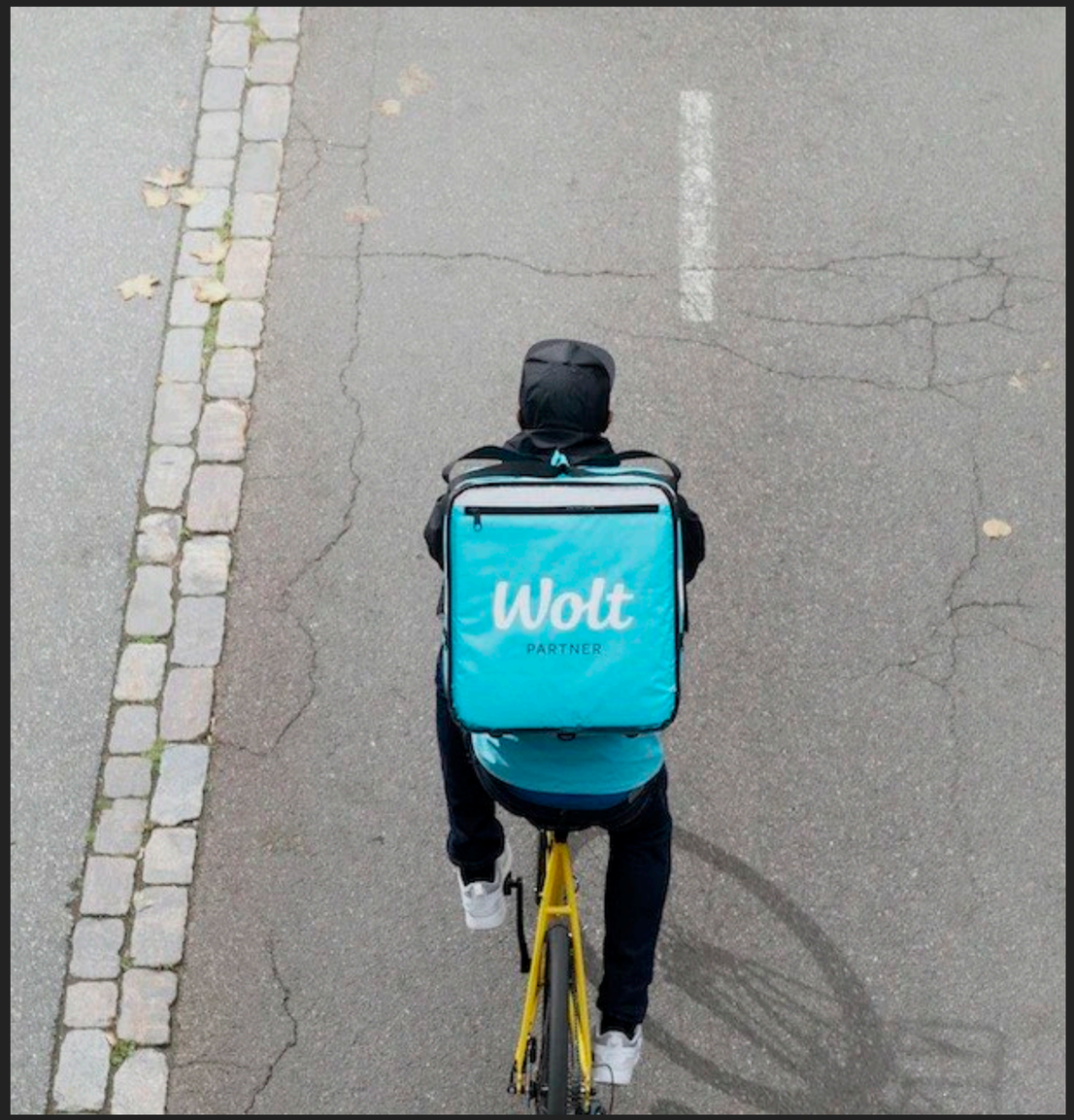
- We have 2 years of daily observations, including the number of courier partners online, average temperature, relative humidity, and precipitation.

THE QUESTION IS

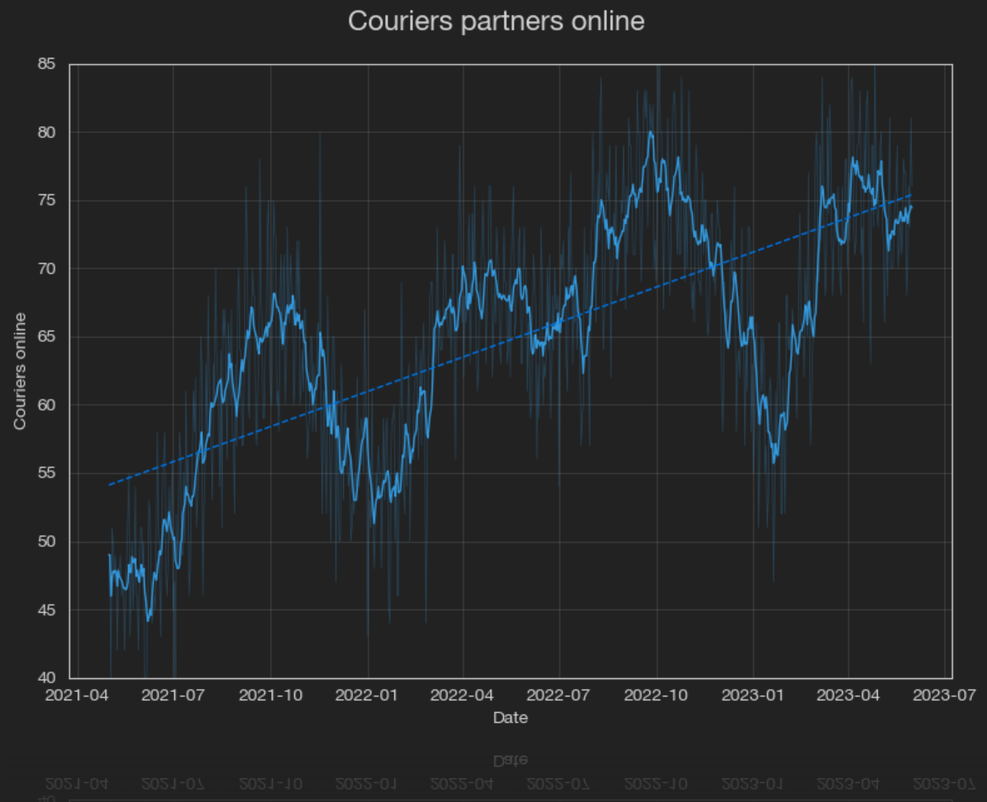
- Based on past data, can we forecast the amount of courier partners that will go online the following day, week, month, or even longer?

LET'S CHECK THIS OUT!

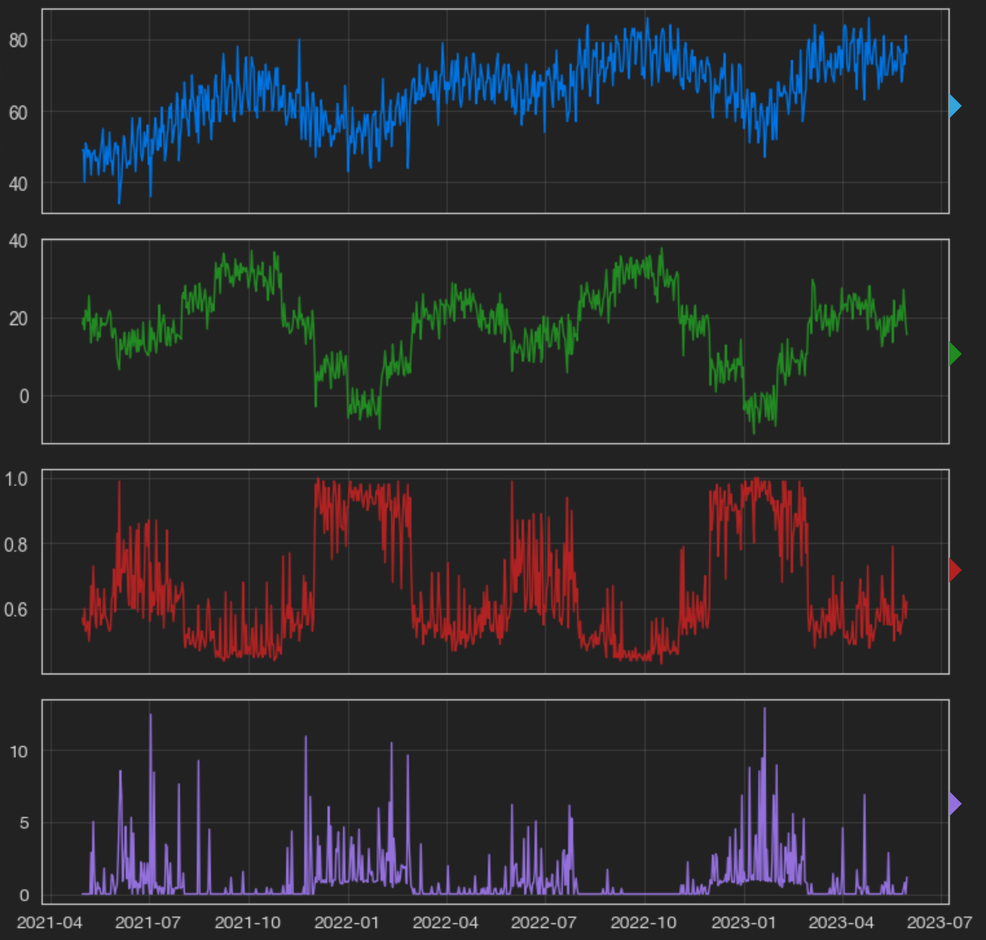
- Spoiler: Surely, we can!



VISUAL REPRESENTATION OF DATA

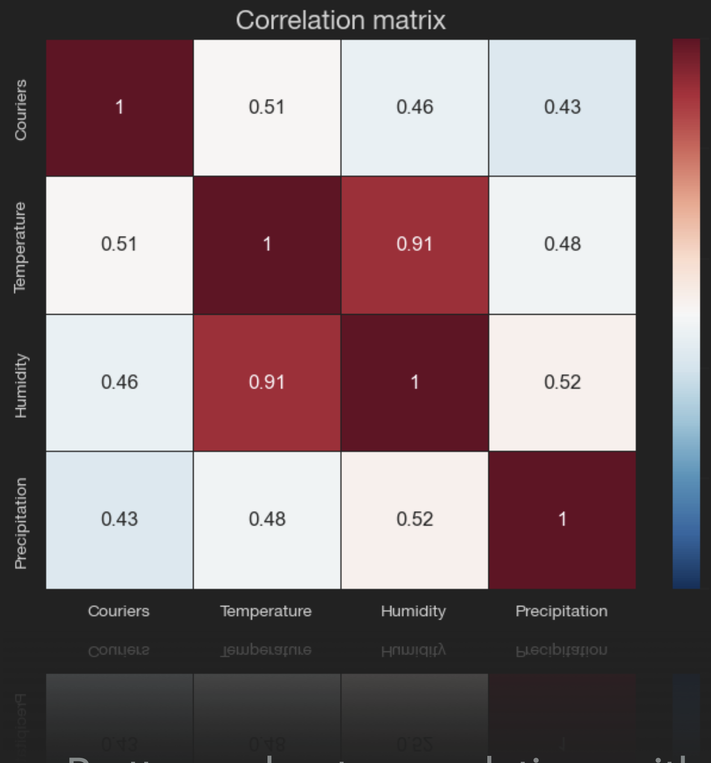


- Main variable
- Seasonality and rising trend
- Obvious relations with the features

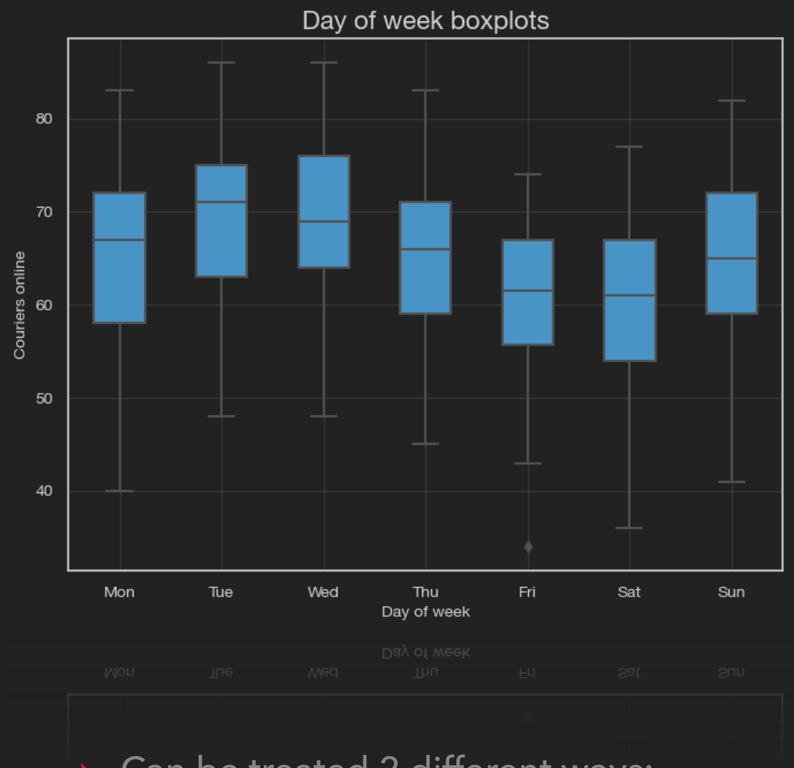


- Couriers online
- Temperature
- Humidity
- Precipitation

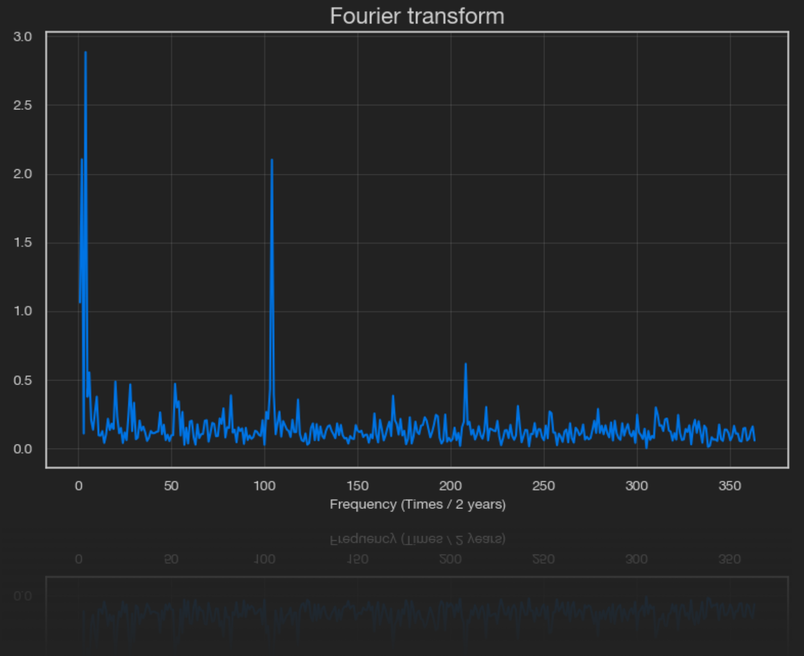
STATISTICS



- Pretty moderate correlations with each feature
- Very high correlation between temperature and humidity



- Can be treated 2 different ways:
- As 3 different groups (ordinal) for a regression model
- As a sinusoid for a seasonal model



- 3 main frequencies discovered, with periods of:
- Week
- 1/2 year
- Year

CONCLUSIONS AND MODELING

- **I see two main approaches for modeling: one involves known (or predicted by meteorologists) weather conditions, while the other does not.**

1. Approach with Known Weather Conditions:
   - For this scenario, a simple linear regression model can be employed. The model would include factors such as the day from the beginning, the day of the week (categorized ordinally), temperature, and precipitation.
2. Approach with Unknown Weather Conditions:
   - In this type of model, reliance on weather condition data is not possible, and the focus is solely on existing time series. Therefore, it makes sense to explore models suitable for time series analysis, particularly those capable of predicting seasonality and trends.

- **Personally, I would consider the following simple model for prediction in the absence of known weather conditions:**

$$y = I + Sx + \sum_{i=1}^{3} A_i sin(2\pi F_i t + P_i)$$

Where I - intercept, S - Slope, A_i - amplitudes, F_i - frequencies, P_i - phase shifts

  - The next step would involve training the model to find the most suitable coefficients.

LINEAR REGRESSION MODEL

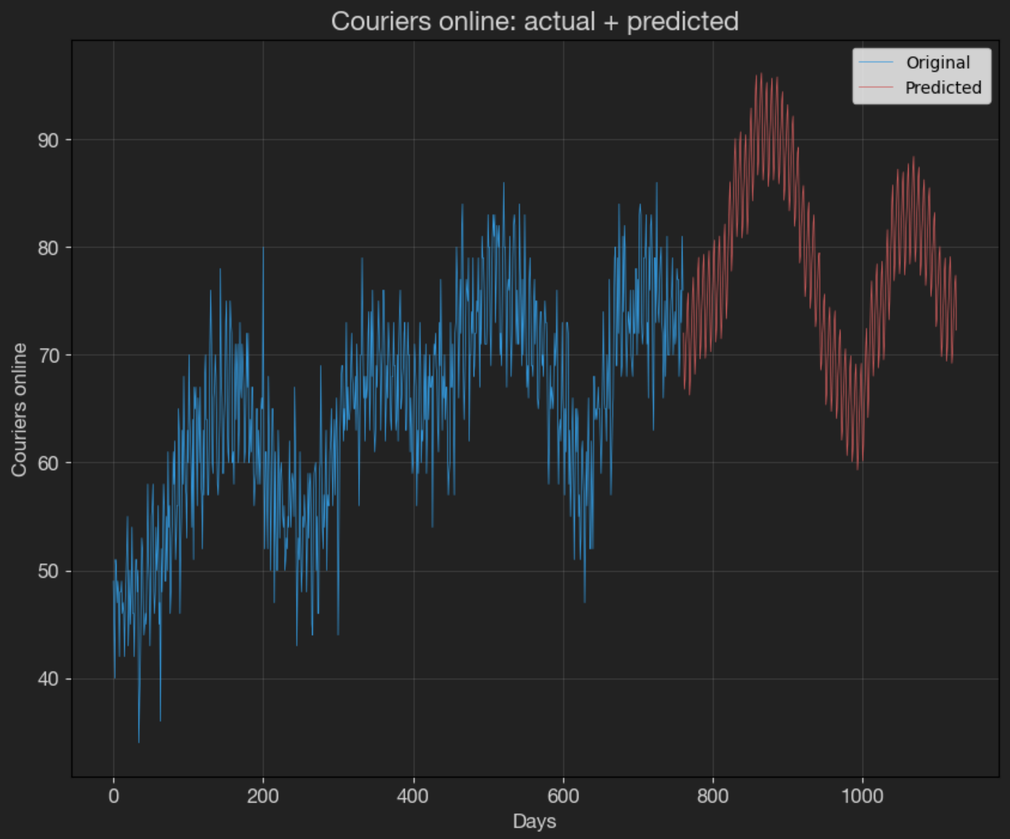
- The prediction curve closely replicates the visual shape of the original curve.
- MAE is 1.68, MSE is 4.56
- Average is 73.6 and amplitude about ±6.
- Mean absolute error 2 times less, than if predict with the average line.

- This model could be useful for short term predictions.

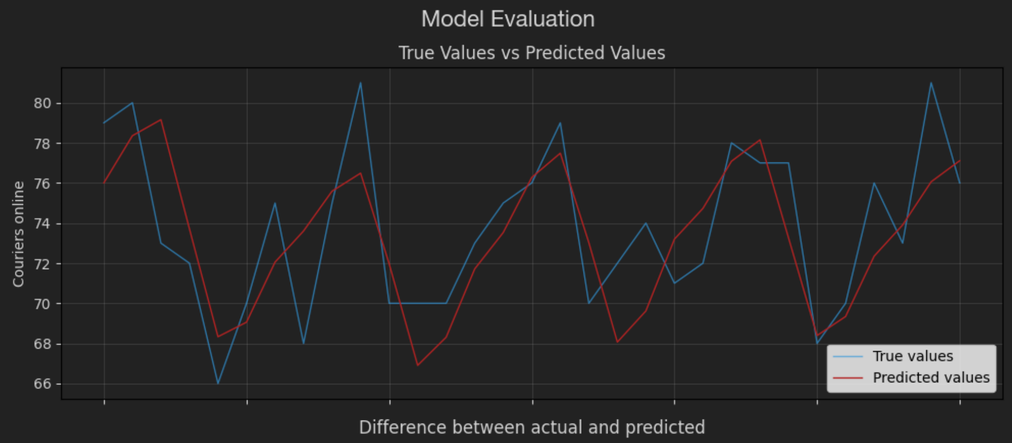
- Strengths:
- Simplicity
- Very easy to tune for the specific time period which better represents the last trends

- Weaknesses:
- Depends on weather forecast

TBATS MODEL



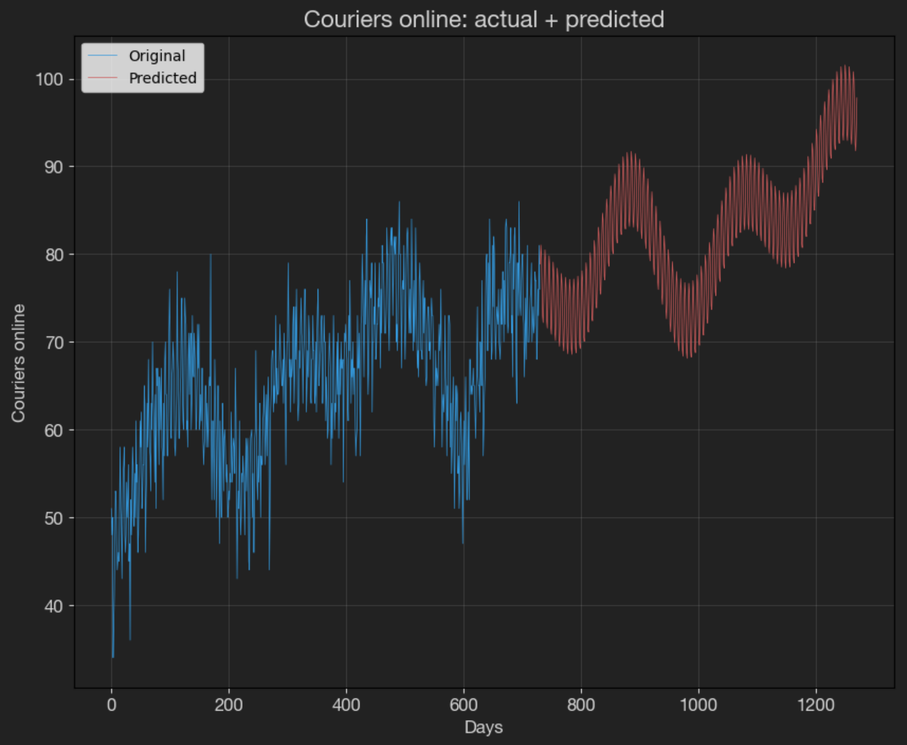
- **Long term prediction**
- The shape of the prediction closely follows the original pattern.
- AIC score is 1096.6



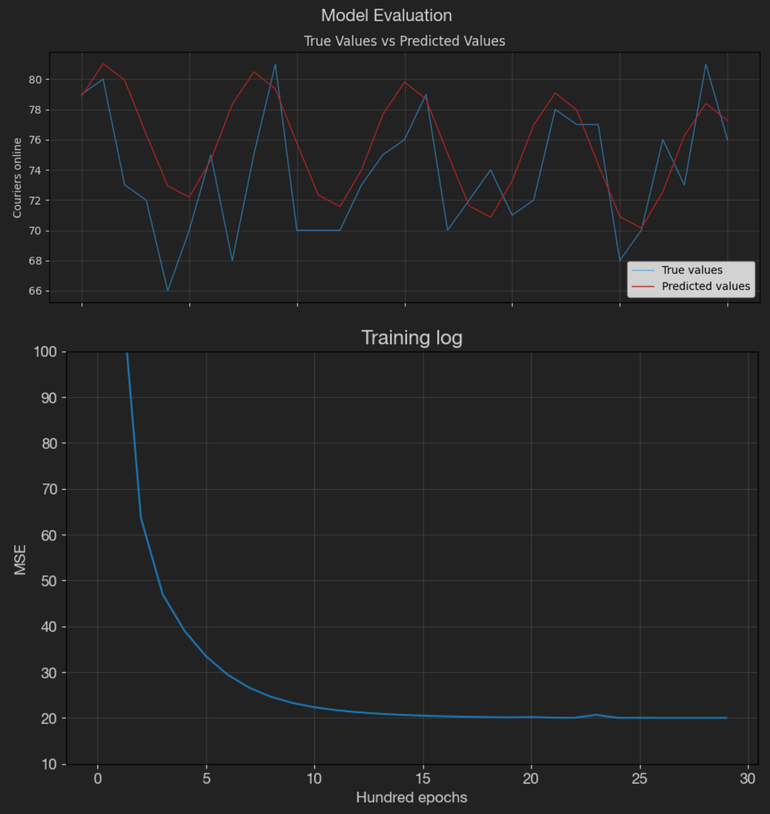
- **Last month prediction**
- The shape is pretty close.
- MSE is 8.22, MAE is 2.4
- Average is 73.6 and amplitude about ±6.

- This model works well for both short term and long term predictions
- Strengths:
- Strong prediction abilities
- Weaknesses:
- Relatively heavy and slow

MY OWN MODEL



- **Long term prediction**
- The shape of the prediction closely follows the original pattern, but less detailed.



- **Learning process took ~1500 epochs with Adam optimizer and learning rate 0.01**
- Overall MSE is 20

- **Last month prediction**
- The shape is relatively close
- MAE is 2.94, MSE is 14.17
- Average is 73.6 and amplitude about ±6.

- This model works well for both short term and long term predictions
- Strengths:
- Good prediction abilities for any term
- Simplicity and clarity, easy to tune
- Weaknesses:
- Restricted by the prescribed equation

WHAT DO WE FINALLY HAVE?

- Applicability:
  - To know the amount of Couriers online could be pretty useful for any planning for money, logistic, time of orders delivery and so on. These models decrease the uncertainty meaningfully and make expectations more reliable.
  - If the data from another places have the same patterns - these models can be used broadly.
- Performance:
  - The best performance is shown by Linear regression model. The only problem is we have to know the weather.
  - For unknown weather TBATS model has the best performance for both short and long term forecasting.
- Complexity:
  - The most complex model is TBATS, where PyTorch and Linear models are the easiest.
- Further development:
  - Any model should be checked and updated from time to time to follow the recent trends. At the same time data for the bigger period must be studied to check some longer patterns.
- For evaluation I see 2 different approaches:
  - Visual observation: How close the prediction curve to the actual one? How well it follows the trends?
  - Comparison of MAE and MSE of model prediction and of just average line of the period. How many times is the model better than just average line? (for short term prediction).
- Further, I would try to add additional parameters to equation in PyTorch model to make better the most simple model. To keep it the most clear model and save computational resources.

THANKS FOR WATCHING!