



# PREDICTIVE TASK

---

Kristijan Rebrović

# OVERVIEW

- Data Location: Kaggle
- Datasets:
  - -Features data set.csv
  - - sales data-set.csv
  - - stores data-set.csv
- Task: predict future sales and conclude

# **STEP 1 – DATA PREPARATION**

- Goal:
- Remove redundant, irrelevant, inaccurate data from datasets
- Check datatypes

## **STEP 2. DATA ANALYSIS AND CLEANING**

- Preprocessing(DQ)
- Markdowns have most missing values, too many to intrapolate
- CMI and Unemployment both have 581 missing values so we intrapolate
- Dates in dataset do not overlap, we need to delete non overlapping dates
- Next we need to merge datasets

# STATISTICAL ANALYSIS

- checking for seasonality in time series – since it's time series, we need to check if there is seasonality pattern
- Correlation – we need to check if variables correlate
- Statistical tests: p-value

# **FITTING**

- Testing if model is fitting the data
- Calculating statistical indicators like Residual Sum of Squares (RSS), Residual standard error (RSE)
- Results show that RSS and RSE are high

# FEATURES RELATIONSHIP WITH TARGET

- T and P values
- If p-value is small, feature is likely to have a relation with target
- In this case we conclude that every feature except "Temperature" has good relationship with target

# **M O D E L I N G**

- Task: Create two separate models and make a comparison
- 1. Linear regression using Prophet
- 2. Exponential smoothing