

# Streaming Uber Taxi Demand Prediction

Liz Aharonian & Ori Ben-Zaken

Supervisors: Barak Bar-Orion, Yohana Khoury, Yoav Einav, Tal Doron



01



## Data collection, cleansing and processing:

Process the data to calculate demand per time and area. Adding relevant features for demand prediction, like: Was the ride on holiday? On weekend? etc.

02



## Training ML Model:

Using Scikit-learn RandomForest model and Spark for supporting distributed environment.

03



## Data streaming:

Using Apache Kafka in order to stream new data in real-time for prediction. Storing the data in Insight-Edge data grid.

04



## Visualization of the data (BI):

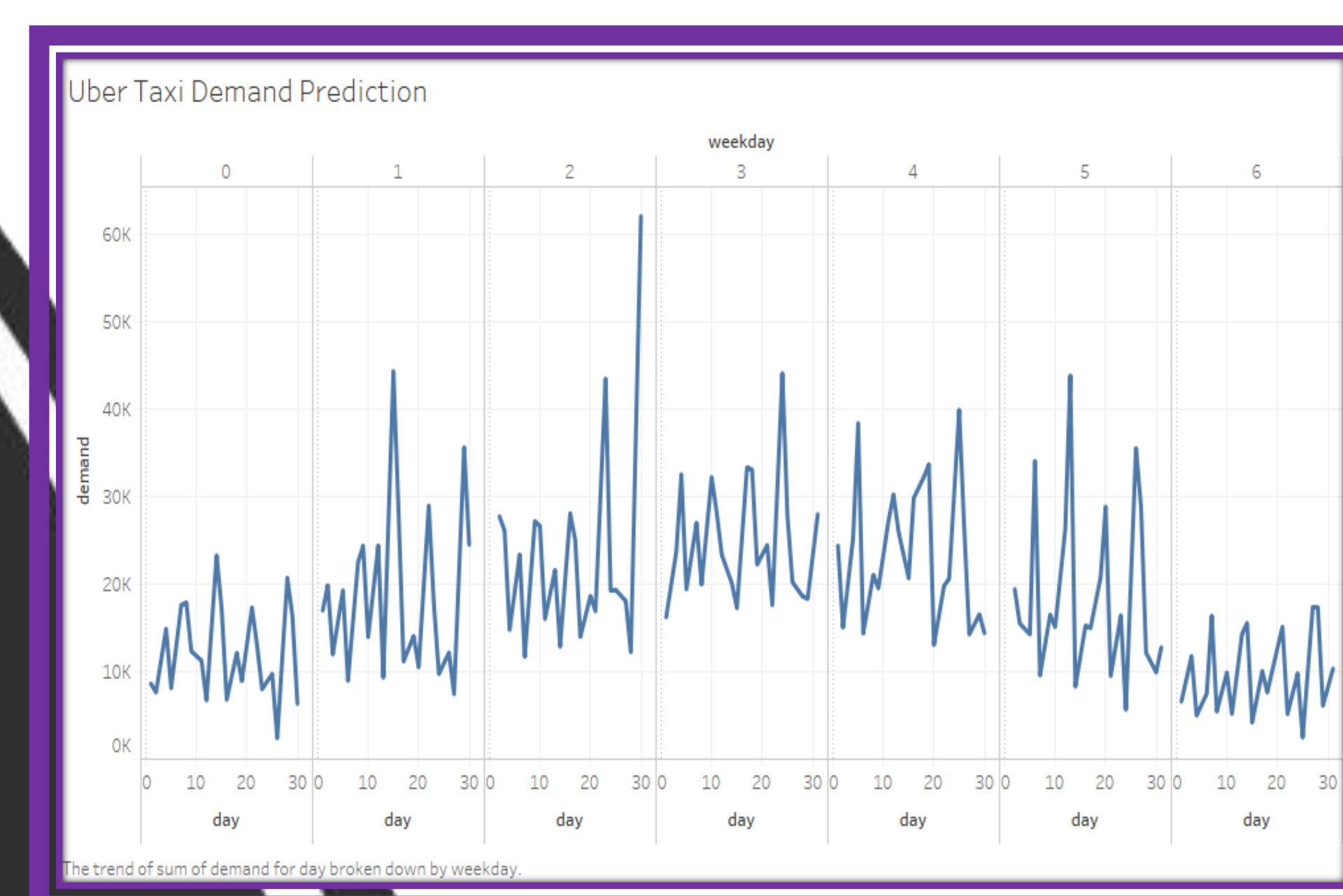
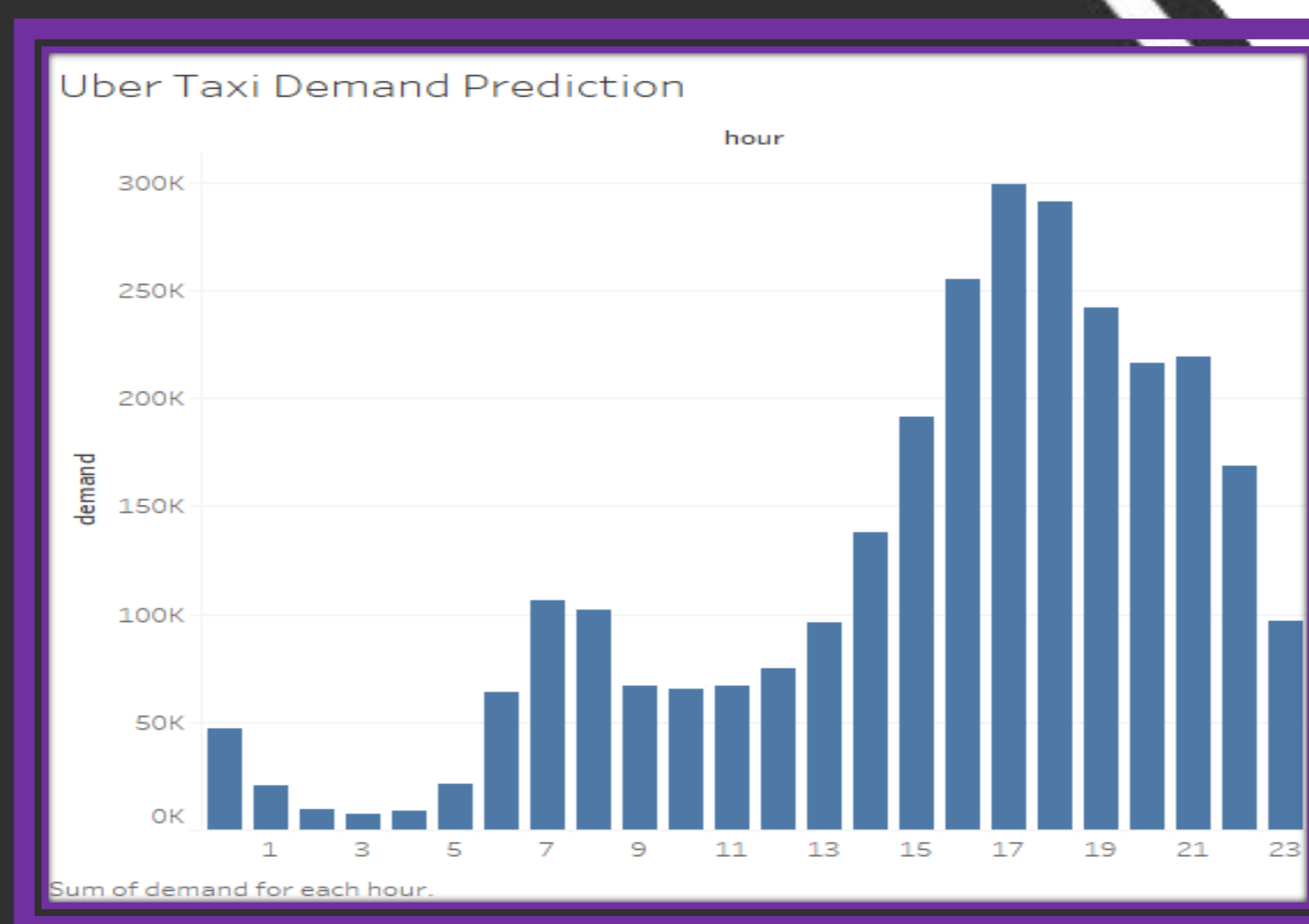
Using Tableau in order to present the model results and perform analytics on the data

### Our goal:

Predict demand per time interval for Uber Taxi Ride, using ML and big data processing tools.

### Demand prediction can:

Help the model's users to know which area they should give service. By using a simple function based on the demand it can calculate the fare amount.



05



## Packaging the model:

Package our model and all its dependencies using Docker technology, a computer program that performs operating-system-level virtualization.