



Bip xTech

Hands-on Challenge

May 2018

Problem Description

Context

- Our customer, one of Europe most important retailers would like to introduce a sales forecasting system to optimize promotions and warehouse stocks.

Objective

- The goal of the project is to provide a working forecasting model and for each store the sum of product sold for March and April.
- The provided training dataset contains daily information about number of sales, number of visitors, store information, weather information and demographic information.

Activities

- Perform a data analysis aimed at analyzing trends and correlation between the available information
- Study and analyze solution/logics for the training set transformation and structuration
- Design and train a forecast prediction model
- Predict the sum of sales for each store for March and April.



Dataset Schema

StoreID	Store ID
Date	Observation Date
IsHoliday	Regional Holiday
IsOpen	Store Open
HasPromotions	The Store is currently running promotions
StoreType	Type of Store (Standard Market, Super Market, Hyper Market, Shopping Centre)
AssortmentType	Type of Products (General, With Fish Department, With Non-Food Department)
NearestCompetitor	Distance (meters) to the nearest competitor store
NumberOfCustomers	Number of Customer
NumberOfSales	Number of Products Sold (<i>Target Value</i>)
Region_AreaKM2	Region Size in which the Store is located (Km2)
Region	Region ID
Region_GDP	Region Gross Domestic Product
Region_PopulationK	Region Population Size (k)
CloudCover	Level of Covering (0-9)
Events	Meteorological event
Max_Dew_PointC	Max Dew Point (Celsius)
Max_Gust_SpeedKm_h	Max Gust Speed (Km/h)
Max_Humidity	Max Humidity
Max_Sea_Level_PressurehPa	Max Pressure (hPa)
Max_TemperatureC	Max Temperature (Celsius)
Max_VisibilityKm	Max Visibility (Km)
Max_Wind_SpeedKm_h	Max Wind Speed (Km/h)
Mean_Dew_PointC	Mean Dew Point (Celsius)
Mean_Humidity	Mean Humidity
Mean_Sea_Level_PressurehPa	Mean Pressure (hPa)
Mean_TemperatureC	Mean Temperature (Celsius)
Mean_VisibilityKm	Mean Visibility (Km)
Mean_Wind_SpeedKm_h	Mean Window Speed (Km/h)
Min_Dew_PointC	Min Dew Point (Celsius)
Min_Humidity	Min Humidity
Min_Sea_Level_PressurehPa	Min Pressure (hPa)
Min_TemperatureC	Min Temperature (Celsius)
Min_Visibilitykm	Min Visibility (Km)
Precipitationmm	Precipitation Amount (mm)
WindDirDegrees	Window Direction (degrees)

Output Format

StoreID	Store ID
Month	Month of the Prediction
NumberOfSales	Number of Product Sold (Prediction)

Evaluation Metric: *minimize the total error*

$$E_r = \frac{\sum_{i \in S_r} \sum_{j \in \{3,4\}} |a_{ij} - p_{ij}|}{\sum_{i \in S_r} \sum_{j \in \{3,4\}}^j a_{ij}}$$

Region Error

$$E = \frac{\sum_{r \in R} E_r}{|R|}$$

Total Error

R Set of regions

a_{ij} Actual Value of Store i and month j p_{ij} Predicted Value of Store i and month j

Dataset Download:

https://drive.google.com/drive/folders/1BGLk_n5qs8xoe8Ahp5s9-1HsQECDvVSB?usp=sharing