Case Study on-

Weather in Szeged 2006-2016

**Objective-**

Using an ML Technique find out the factors having an impact on the Temperature.

**Dataset Variables**-

* time
* summary
* precipType
* temperature
* apparentTemperature
* humidity
* windSpeed
* windBearing
* visibility
* loudCover
* pressure

**EDA-**

**Rows and Columns**: 96453 rows × 12 columns

**Duplication**: Found 24 duplicated rows which was removed.

**Null Values**: 517 null values for “Precip Type” variable which is 0.53% so removing such

small observations won’t make much difference so dropped all null values.

**Irrelevant variables**: ("Formatted Date”,” Loud Cover”, “Daily Summary") which doesn’t

impact our predictions.

**Outliers:** All variables mean and median is almost similar so there is no outlier in any

variables.

**Visualization-**

**When “Precip type” is rain** -Most of the time the Temperature is between-16-20

degree(C), Humidity between 0.9-1.0,Wind speed in the

range 6-12 km/hr,Visibility 11-13 km and Pressure

between 950-1050 milibars

**When “Precip type “ is snow**- Most of the time Temperature is between -3 to 0 degree

(C),Humidity between 0.9-1.0,Wind speed in the range 5-

10 km/hr,Visibility 10-12 km and Pressure between 900-

1050 milibars.

**Effect of variables on dependent variables which is Temperature**-

-As humidity increases till 0.1 temperature **increases** then with humidity

temperature keeps on decreasing.

-Wind speed, wind bearings and Pressure **doesnot have much impact** on

temperature as mean of temperature is constant throughout.

-With **increase in visibility** there is very minute increase in temperature.

-From heatmap we can find **there is correlation between** few independent

variables.(Pressure,visibility,humidity)

**Feature Engineering:**

-Converting categorical features into numeric.

-Under “Summary” variable category which have less than 10% of total

Observations categorized as Others and rest as per the original category.

-Creating dummy variable of variable “Summary” and “Precip Type”.

-Dropping one dummy variable of “Summary” category and one variable of

“Precip Type” category to avoid multicollinearity and dummy variable trap.

**Creating the model:**

- Call linear Regression.

-Divided the data into 70-30 ratio of train and test data .

-Fit the model.

-Found values of intercept and coefficients.

-Model predicted.

**Model Evaluation:**

**Train model accuracy**-98.97%

**Test model accuracy**-98.99%

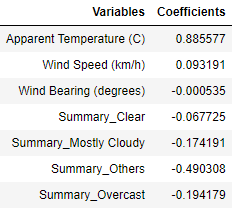
**Overall model accuracy**-

R2 :0.9899325620861553

MAE :0.7514411035401882

MSE :0.927433308818046

MAPE :16682753084448.756



-So as Apparent temperature increases by 1 unit Temperature will increase by 0.88 units.

- as Wind Speed increases by 1 unit Temperature will increase by 0.093 units.

-as Wind bearing (degrees) increases by 1 unit Temperature will decrease by 0.00053 units.

- When climate summary which is clear increases by 1 unit Temperature will decrease by 0.06 units.

When climate summary which is Mostly Cloudy increases by 1 unit Temperature will decrease by 0.17unit

When climate summary which is others increases by 1 unit Temperature will decrease by 0.49 units.

When climate summary which is overcast increases by 1 unit Temperature will decrease by 0.19 units.