FORECASTING TEMPERATURE USING WEATHER TIME SERIES DATA OF US & CANADIAN CITIES FOR EFFECTIVE INVENTORY STOCK PLAN AT RETAIL STORES

Team B10

Rahul Sharma
Himanshu Jain
Rahul Garg
Artika
Vividha Arabandi
Varghese Zachariah

Business Goal

Forecasting of temperature for inventory stock planning in Retail Stores

We have collected data containing ~5 years of high temporal resolution (hourly measurements) data of weather conditions in 3 US cities and 3 Canadian cities **(Chicago, Denver, Detroit, Vancouver, Toronto, Montreal).**

- The business goal we have chosen for this assignment is the forecasting of temperature so that retail owners like **Walmart, Kmart, Weston and Costco** along with various mom-pop retails can plan their inventory stock up of various crucial SKUs. As the temperature plummets to freezing zone, inventory replenishment becomes harder and costlier.
- There is also a possible correlation between temperature and vehicular traffic, thus having a direct impact on inventory stock up costs.
- Extreme low temperatures have the potential to cause millions of dollars of overhead expenses and lost sales for the retailers especially in the areas where temperatures are known to dip drastically.

Forecasting Goal

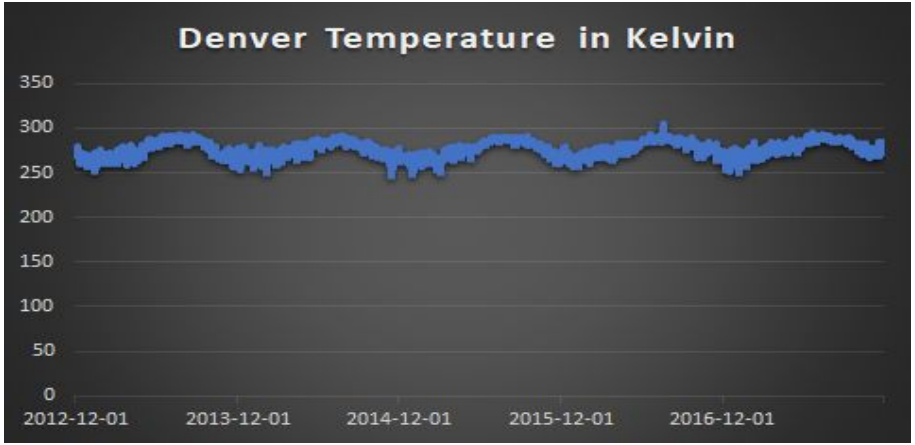
- The data for temperature is from 2012 to 2017, ie the data is quite recent and can be used for forecasting the same.
- Because the business will be interested in the minimum temperature forecast of a day, we have aggregated the data on daily basis - minimum temperature.
- The forecasting time scale would be on daily basis.
- We will forecast temperature for 6 US and Canadian cities mentioned above for a period of one year.

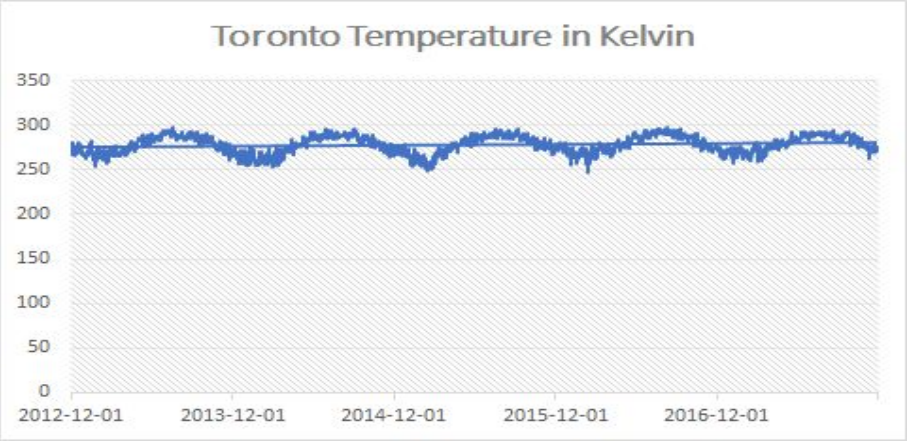
Data

Data source: Kaggle
(https://www.kaggle.com/selfishgene/historical-hourly-weather-data#temperature.csv)

Data Attributes
Timestamp | City name | Temperature in Kelvin

Time Series Components:
Level
Noise
Trend - Linear (very flat)
Seasonality





Data Preparation

Step-1: Missing Value Treatment - Used Mode

Step-2: Change of Scale : Daily Aggregation from hourly - Used min temp

Model Preparation

Step-1: Added columns ‘t’ , ‘t^2’ , and created Categorical variables for linear regression for Season index

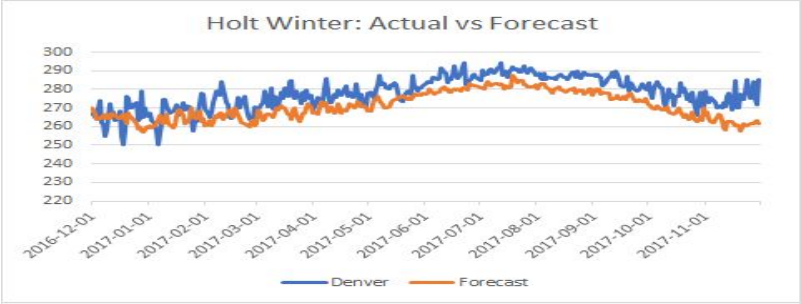
Step-2: Data Partitioning

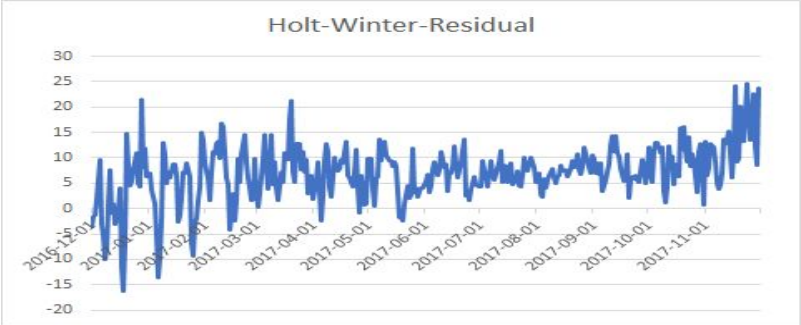
Training: 4 years (1461 records) Validation: 1 year (365 records)

Benchmarking

We have used Naive forecast (lag 365 value) for benchmarking

Method-1: Smoothing by Holt Winter (Triple Exponential) with additive seasonality





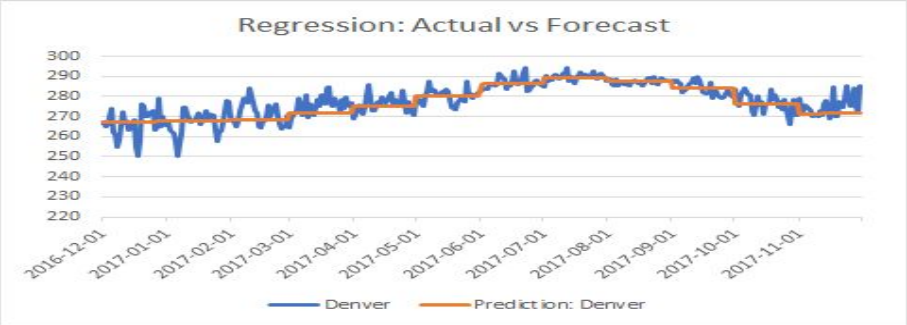
Optimize Params	Yes
Period	365
Alpha (Level)	0.584918
Beta (Trend)	0.003815
Gamma (seasonality)	0.008179

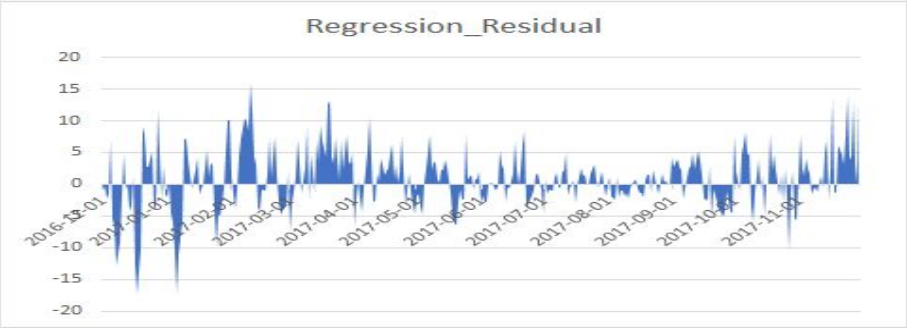
Record ID	Value
SSE	28687.15
MSE	78.59492
MAPE	2.777486
MAD	7.751918

RMSE	8.8653777

Trend is captured

Method-2: Regression with linear trend



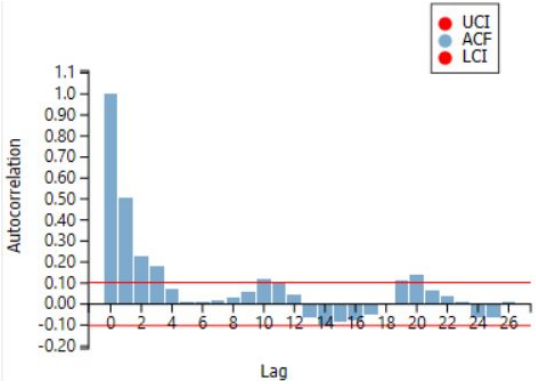


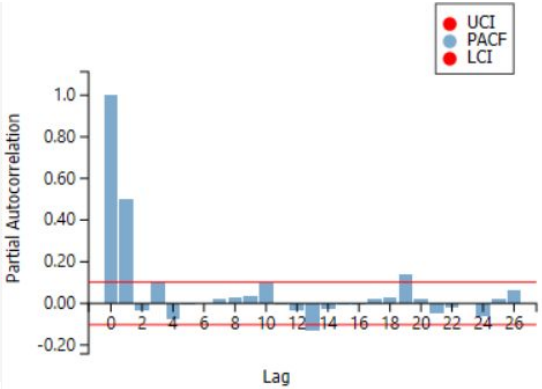
Metric	Value
SSE	7883.837
MSE	21.59955
RMSE	4.647532
MAD	3.440248

RMSE	4.647532

Seasonality is captured

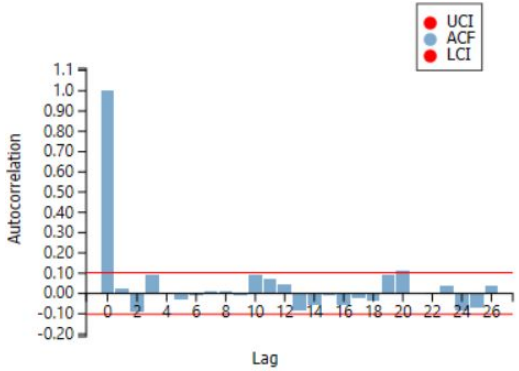
Autocorrelation Check in Linear Regression Model

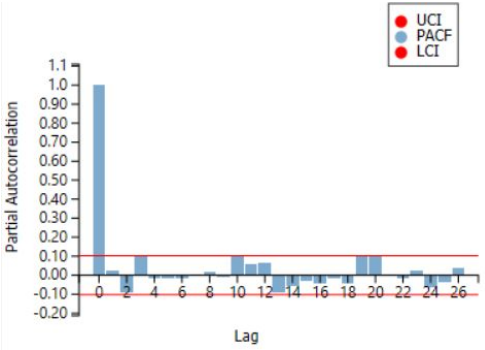


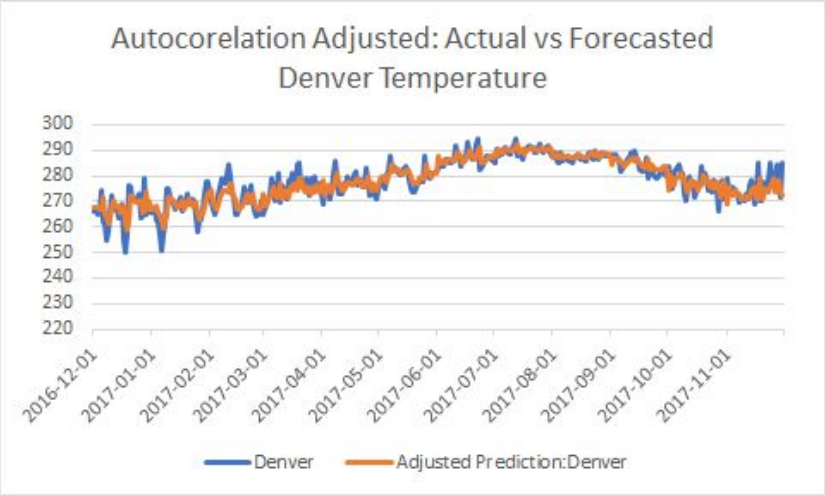


ARIMA

Record ID	Coeff
Const	0.401557
AR 1	0.513114

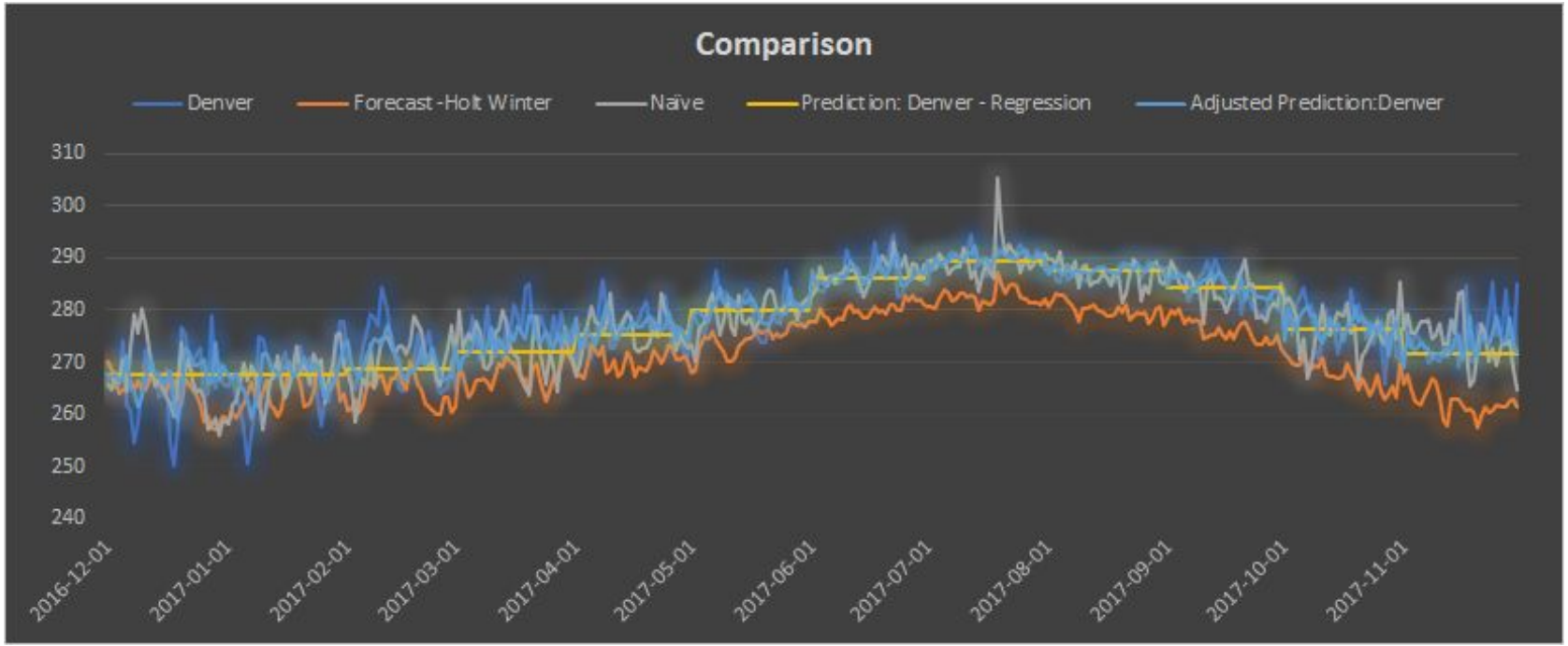






Model Comparison

Method/Metric	Naïve	Holt-Winter	Linear Regression	Regression with Autocorrelation adjusted
RMSE	6.019471166	8.865377735	4.647532038	2.45707489



Recommendations

Learning:

~While Linear Regression performed the best, it was a staggered forecast.Autocorrelation check is important.

~Short forecast is better as less error from extreme weather pattern or anomaly.

~Use Control Charts for model performance. Review and adjust the model after a fixed period (every 6 months) or when control chart indication of error.

~The MAPE assumes that percentages make sense; that is, that the zero on the scale of the data is meaningful. But when forecasting temperatures in degrees Celsius or Fahrenheit it makes no sense. The zero on these temperature scales is relatively arbitrary, and so percentages are meaningless.

Deployment:

Retail Stores can use this model for forecast weekly temperatures.
They can stock inventories on the day of relatively highest minimum temperature.
This model can also deployed for hourly forecasts for retailers who do daily stock-ups of inventory