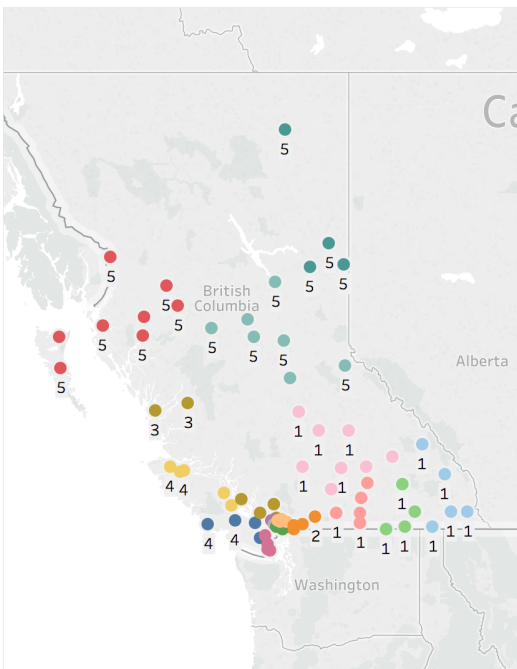Two dashboards to be shown are made by Tableau Desktop for mac.
The csv formatted data were collected from the websites, city of Vancouver and Government of British Columbia, in order to demonstrate the project management skills and data analysis skills.
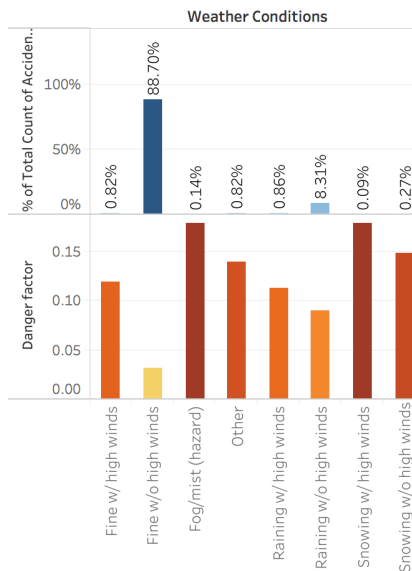I can also construct warehouse using AWS S3, redshift, EC2 and Tableau to make scalable data analysis.

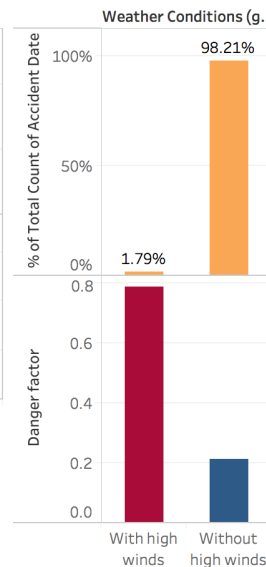## Namshik's Tableau Portfolio

hopital locations

HLTH_SDA_N
- Central Vancou..
- East Kootenay
- Fraser East
- Fraser North
- Fraser South
- Kootenay Boun..
- North Shore / C..
- North Vancouve..
- Northeast
- Northern Interi..
- Northwest
- Okanagan
- Provincial Healt..
- Richmond
- South Vancouve..
- Thompson Cari..
- Vancouver

Correlation of accidents and wether

Weather Conditions

% of Total Count of Acciden..: 0.82%, 88.70%, 0.14%, 0.82%, 0.86%, 8.31%, 0.09%, 0.27%

Danger factor — Fine w/ high winds, Fine w/o high winds, Fog/mist (hazard), Other, Raining w/ high winds, Raining w/o high winds, Snowing w/ high winds, Snowing w/o high winds

Grouped visualization

Weather Conditions (g..

% of Total Count of Accident Date: 1.79%, 98.21%

Danger factor — With high winds, Without high winds

time series forecasting

Month 2019 — January / February

measured v..: 266.0, 231.8, 336.5, 273.3, 287.0, 264.5, 315.9, 440.4, 575.5, 581.3, 475.3, 601.3, 724.1, 748.7, 797.9, 871.6

The left map shows the locations and categorizations for Hospitals in British Columbia. Using the address or longitude/latitude data, we can visualize the clustering very effectively.

The two right upper histograms is to analyze the correlation of the traffic accidents and weather conditions. The majority of accidents occur in fine days without high winds by data and simple management(Blue histograms). However, when use a **calculated fields** function and make a newly defined rows danger factors by proper syntax, it is analyzed when it is extremely foggy or snowy with high winds, people should drive with the most caution. When grouping the columns by winds, the high windy days are almost 4 times more dangerous than days without high winds.

Tableau also supports time series forecasting shown in the right lower line graph. The light blue line and bands present the expected graph and 95% confidence interval. I can also utilize the time series forecasting using machine learning and statistical models by Python.
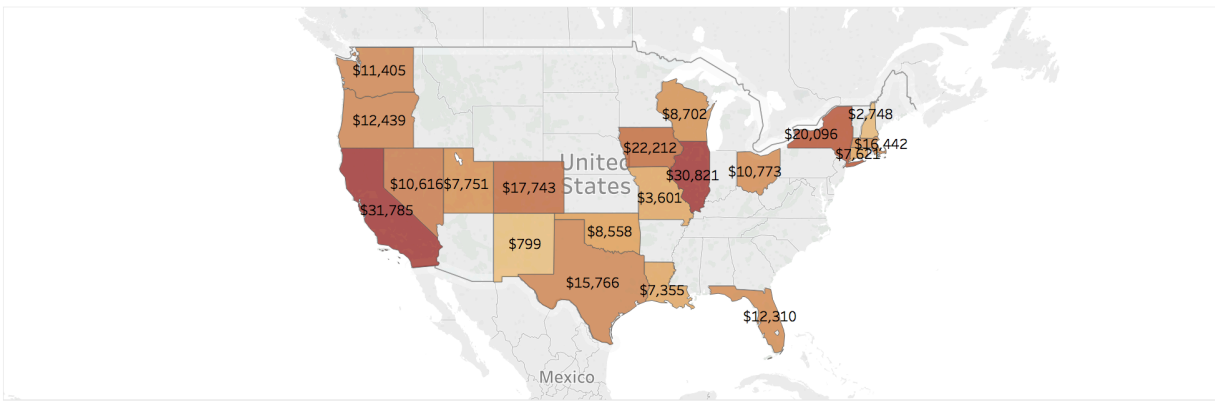
For example, the time series forecasting method used above is the exponential smoothing with additive trend and without seasonality. The trend and seasonality were chosen automatically by Tableau itself. However, it doesn't show the detailed process including loss of the training. Besides, the multiplicative trend yields much broader confidence interval and the interval is growing as time evolves.

My data visualization skill is not restricted in Tableau, but I am also proficient at various and useful Python libraries for data analysis, optimization, prediction and visualization. The exponential smoothing forecasting by Tableau can be generalized to autoregressive integrated moving average (ARIMA) model with specific 3 integer parameters, (p,d,q). ARIMA model is a powerful tool in time series analysis and forecasting. The following forecasting and the 95% confidence interval is predicted by the ARIMA model with (p,d,q) = (4,0,0). The parameters are chosen by minimum loss optimization. I have used the statsmodel, Pandas, matplotlib, sklearn, numpy library in Python, in order to implement the ARIMA model. The data were given sinusoidally and exponentially. The randomness of the model is not quite big, so the confidence interval is quite narrow. The predictions(blue) is very such close to the original values(green line) in the first plot. I can also use LSTM deep neural network learning to forecast the time series forecast.
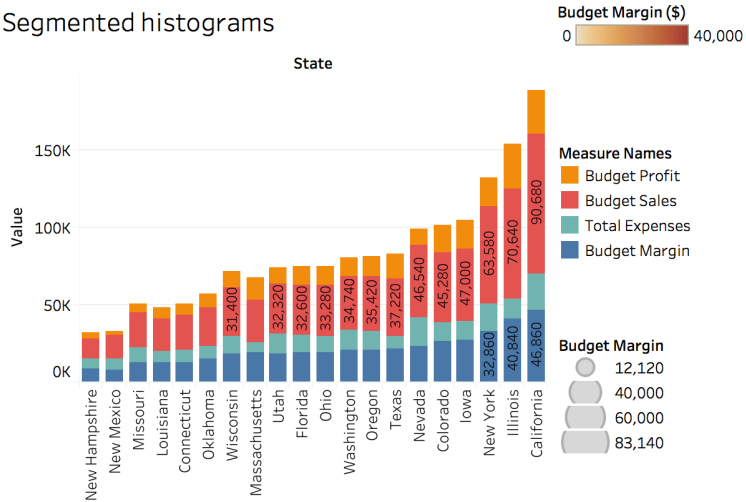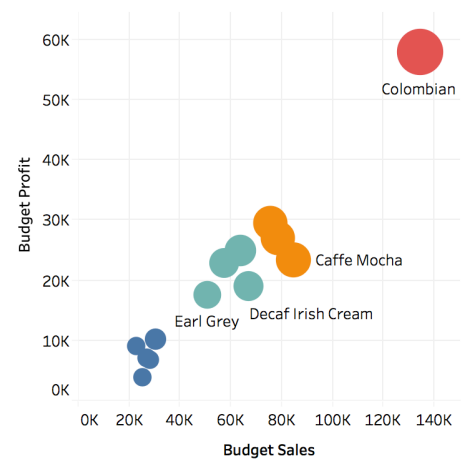

Test MSE: 6.329

Profits and margins at 20 states

Map labels: $11,405 · $12,439 · $8,702 · $2,748 · $20,096 · $16,442 · $22,212 · $7,621 · $10,616 $7,751 · $17,743 · $30,821 · $10,773 · $31,785 · $3,601 · $799 · $8,558 · $15,766 · $7,355 · $12,310

United States · Mexico

Budget Margin ($)
0 — 40,000

Segmented histograms

State

Value: 150K · 100K · 50K · 0K

States (x-axis): New Hampshire, New Mexico, Missouri, Louisiana, Connecticut, Oklahoma, Wisconsin, Massachusetts, Utah, Florida, Ohio, Washington, Oregon, Texas, Nevada, Colorado, Iowa, New York, Illinois, California

Bar labels: 31,400 · 32,320 · 32,600 · 33,280 · 34,740 · 35,420 · 37,220 · 46,540 · 45,280 · 47,000 · 32,860 · 63,580 · 40,840 · 70,640 · 46,860 · 90,680

Measure Names
- Budget Profit
- Budget Sales
- Total Expenses
- Budget Margin

Budget Margin
- 12,120
- 40,000
- 60,000
- 83,140

Clustering

Budget Profit (y-axis): 60K · 50K · 40K · 30K · 20K · 10K · 0K

Budget Sales (x-axis): 0K · 20K · 40K · 60K · 80K · 100K · 120K · 140K

Cluster labels: Colombian · Caffe Mocha · Decaf Irish Cream · Earl Grey

This dashboard shows two visualizations of data and a clustering analysis. The map shows the profits and margins at twenty states in US. The marks in the map shows the profits and the densities of colours present the budget margins by coffee sales. The left lower segmented graph well visualizes the magnitudes of the profits, sales, expenses and margins. The histogram are sorted out by margins. The right lower one analyze what kinds of coffees make bigger profits and sales. Colombian coffee are exceptionally beneficial in the business.

Data scientist
PhD Namshik Kim