PROPOSAL OUTLINE

KEY DISCUSSION POINTS

Plan For The Weather

Does It Even Work?

Only God Knows The Weather

The Roll Out

Done And Dusted

Future Work

Risky Behaviour

Meet The Team

PLAN FOR THE WEATHER

STRATEGY

What does 10°C and light rain mean for sales?
The answer is different in Vancouver than in Ottawa; it varies if it is a day in May or November. Weather's influence on consumers is complex and nuanced; temperature or precipitation data is easily misinterpreted and misused.

Our goal is to demystify these weather myths that affect retailers –
- Weather has no impact on the sales.
- Sometimes the weather hurts me, sometimes the weather helps me. It all evens out in the end.
- Consumers will shop during the holidays, regardless the weather.
- If the product is seasonal, only then the weather would affect it.

HOW DO WE HELP?

Our goal is to analyze the impact of weather and optimize the staffing of stores appropriately to manage the influx of shoppers.

DOES IT EVEN WORK?

WEATHER MATTERS

Forecasting demand in the fashion industry is a notoriously difficult problem. Adding factors such as weather make it even more complicated.

Incorporating weather information into the forecast and ordering decision making processes could help respond faster and potentially benefit a lot. The impact of weather is multi-faceted: less walk-ins due to harsh weather leading to drop in sales or the increased demand for some products and potential increase in profit.

+1.7%

-2.1%

+3.2%

ONLY GOD KNOWS THE WEATHER

HOW FAR CAN WE PREDICT?

Weather is a notoriously fickle and uncontrollable factor, and no forecaster can reliably predict the temperature beyond the next few weeks.

One reasonable approach is to predict retail sales for the upcoming week. Since the weather forecast can be fairly accurate for a week, we can provide product-sales estimates over the course of that week.

THE ROLL OUT

PROPOSED TIMELINE

DATA COLLECTION

Collecting weather data through APIs, for each city and the complete timeframe of the transaction records

EXPLORATORY DATA ANALYSIS

Performing EDA on the existing dataset, and analyzing the trends using timeseries

FEATURE ENGINEERING

Generating useful features from the raw weather data gathered

MODEL CREATION

Using Statistical as well as Deep Learning models to predict short-term product sales

DONE AND DUSTED
DATA COLLECTION

Task: Collect weather data for the sales dataset

API Used: Meteostat

Approach:

- We were provided with the retail-sales dataset consisting of 44 cities and 3 provinces
- Created a mapping of city names and city codes used by the API along with the earliest and the latest dates of transactions for each city
- Received JSON responses from the API for each city and the date range, which were then saved as CSV files

weather_daily

date	temperature	temperature_min	temperature_max	precipitation	snowfall	snowdepth	winddirection	windspeed	peakgust	sunshine	pressure	station_id	station_name
2014-01-01	-11.8	-16.6	-3.0					7.5			1028.3	71872	medicine hat
2014-01-02	-3.0	-18.0	5.0				206.0	19.3			1011.6	71872	medicine hat
2014-01-03	-3.3	-11.3	5.0				302.0	23.7			1006.8	71872	medicine hat
2014-01-04	-17.9	-22.0	-12.0				287.0	18.2			1032.9	71872	medicine hat
2014-01-05	-26.5	-31.0	-21.0				221.0	15.3			1046.4	71872	medicine hat
2014-01-06	-15.0	-25.0	-8.0					10.3			1026.1	71872	medicine hat
2014-01-07	-9.1	-13.0	-5.6					10.3			1020.7	71872	medicine hat
2014-01-09	0.5	-4.0	4.0				228.0	23.0			999.9	71872	medicine hat
2014-01-10	-3.7	-10.0	2.0				235.0	14.1			1002.5	71872	medicine hat
2014-01-11	0.7	-11.0	7.0					25.2			993.1	71872	medicine hat
2014-01-12	-1.5	-6.0	2.0				215.0	17.3			999.4	71872	medicine hat
2014-01-13	2.0	-1.0	6.0				269.0	33.0			1007.3	71872	medicine hat
2014-01-14	0.6	-4.0	4.1				243.0	23.9			1024.7	71872	medicine hat
2014-01-15	4.4	-0.3	8.0				273.0	33.8			1015.4	71872	medicine hat
2014-01-16	0.6	-2.0	4.0				238.0	17.0			1027.3	71872	medicine hat
2014-01-17	1.6	-2.0	7.0					15.8			1019.6	71872	medicine hat
2014-01-18	1.9	-3.0	7.0					18.4			1016.9	71872	medicine hat
2014-01-19	1.6	-1.0	6.0				231.0	15.2			1014.9	71872	medicine hat
2014-01-20	-1.9	-7.0	3.0					11.0			1027.3	71872	medicine hat
2014-01-21	-1.7	-4.2	2.0				260.0	11.1			1027.2	71872	medicine hat
2014-01-22	-8.9	-11.2	-5.0				95.0	11.5			1032.7	71872	medicine hat
2014-01-23	-1.7	-10.0	6.0				224.0	21.0			1027.7	71872	medicine hat
2014-01-24	4.5	1.0	10.0				263.0	20.8			1021.9	71872	medicine hat
2014-01-25	4.6	1.0	10.0				259.0	24.9			1016.3	71872	medicine hat
2014-01-26	-8.1	-15.6	4.0					21.8			1029.2	71872	medicine hat
2014-01-27	-15.4	-18.0	-12.0					4.9			1035.2	71872	medicine hat
2014-01-28	-7.7	-17.0	1.0					13.8			1019.3	71872	medicine hat
2014-01-29	-7.8	-11.4	-4.0					13.0			1009.5	71872	medicine hat
2014-01-31	-14.4	-19.0	-10.0				274.0	21.1			1016.8	71872	medicine hat
2014-02-01	-12.7	-15.7	-10.0				225.0	11.6			1022.6	71872	medicine hat
2014-02-02	-11.8	-18.0	-9.0				269.0	13.3			1022.9	71872	medicine hat
2014-02-04	-20.7	-23.3	-19.0				278.0	12.8			1043.5	71872	medicine hat
2014-02-05	-22.2	-27.9	-18.0					9.7			1047.6	71872	medicine hat

DONE AND DUSTED

EXPLORATORY DATA ANALYSIS

Task: Perform initial investigations on data so as to discover patterns, spot anomalies and check assumptions using graphical representations
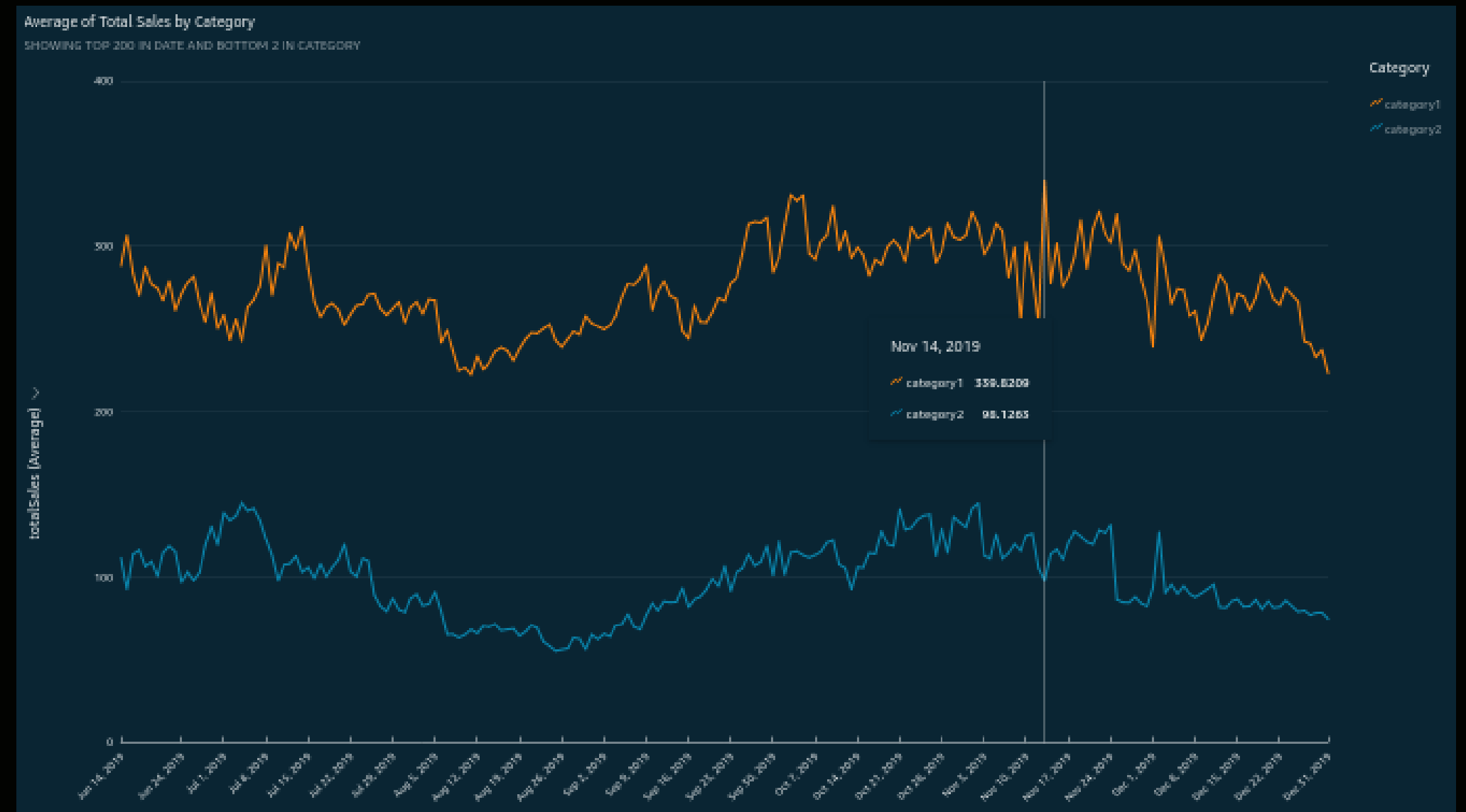
Tools Used: Plotly, Seaborn

Approach:

- Aggregated data over multiple combinations of Date, Province, Category, and other features
- Created timeseries plots from the data generated.
- Initial findings without the addition of weather data show that there are more sales in summer as compared to spring or fall. Detailed analysis will be done at a later stage.

EXPLORATORY DATA ANALYSIS

One such analysis is a TimeSeries plot of sales of Category 1 and Category 2.

Since the distribution of both the plots is different, a single model might not be able to generalize for both the categories.
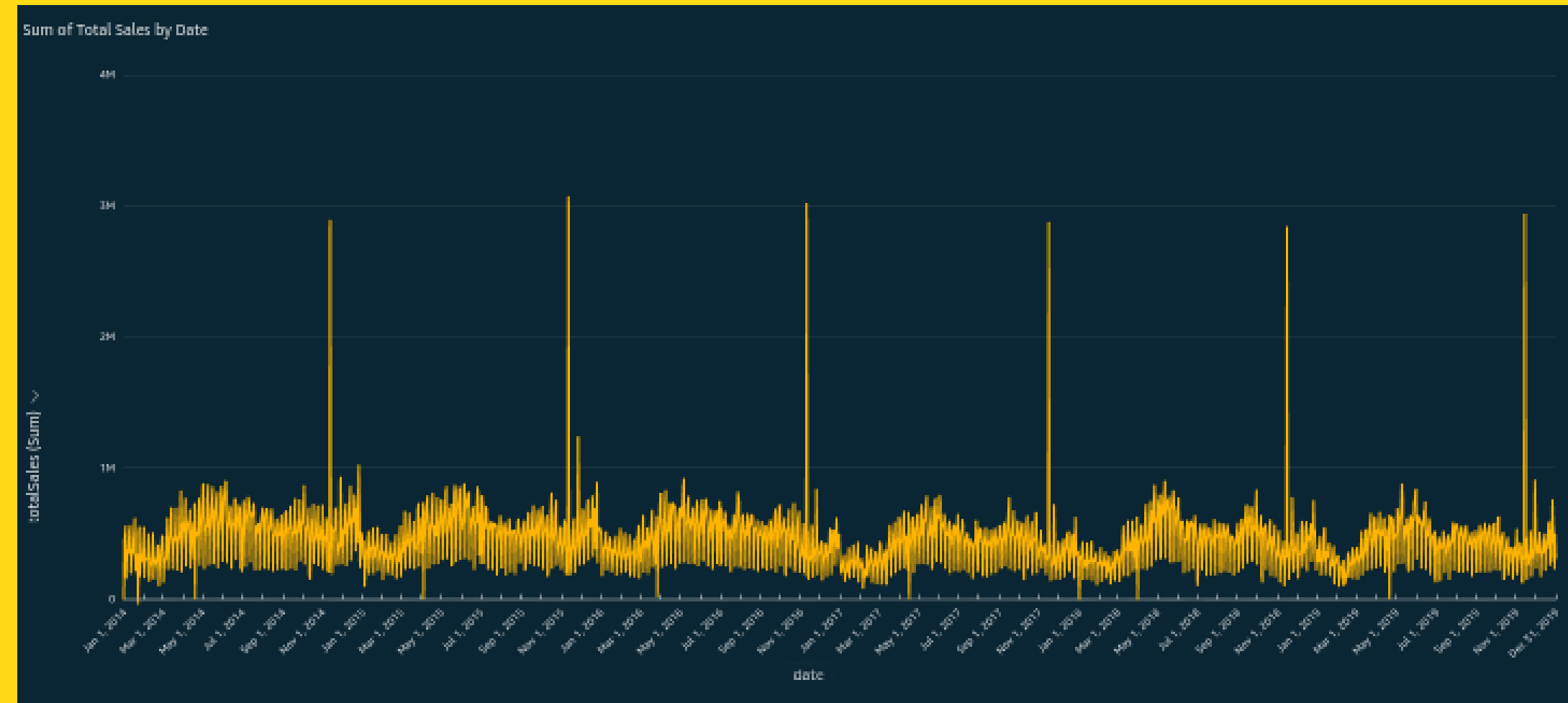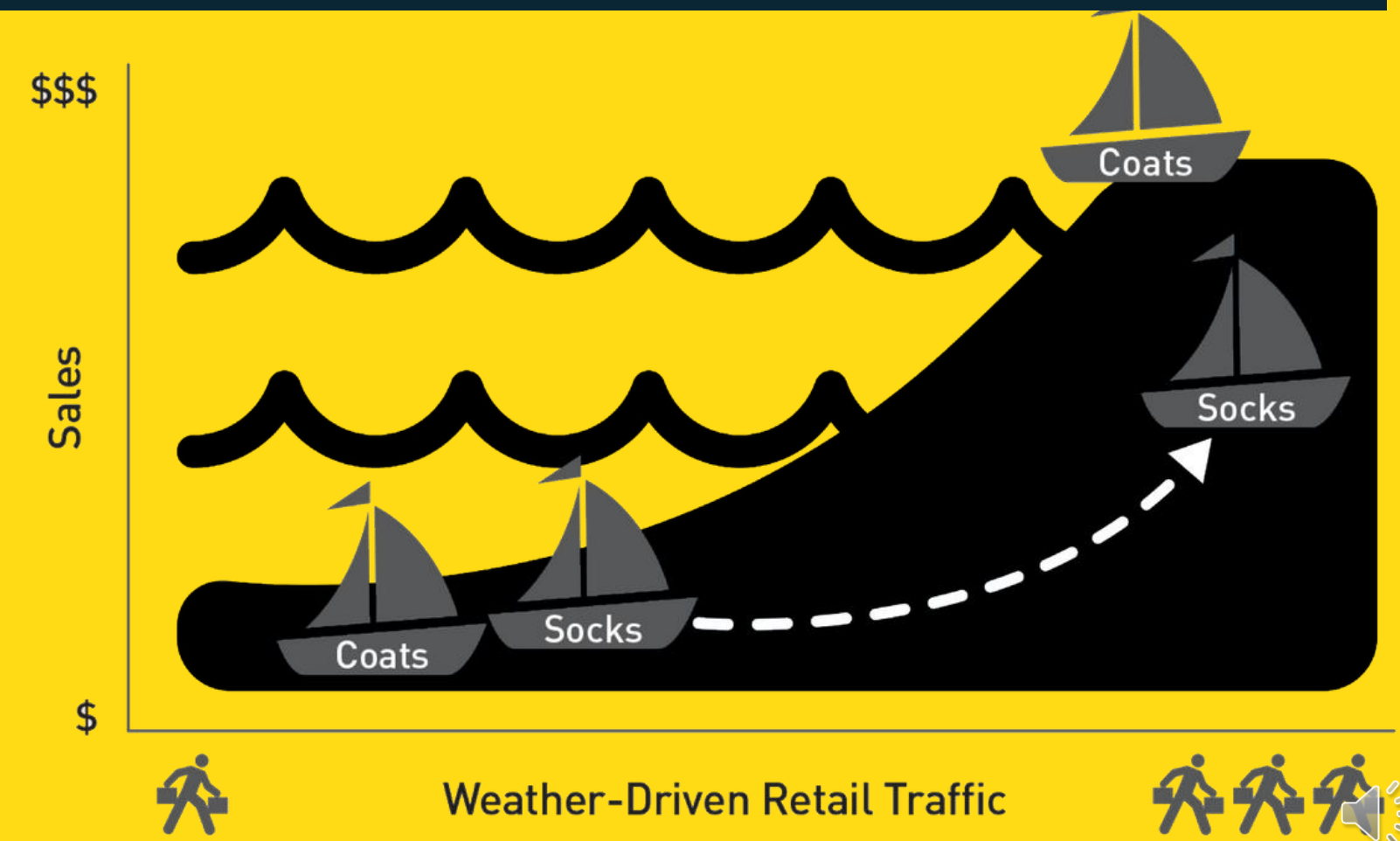
FUTURE WORK

FEATURE ENGINEERING

Task: Generate and select useful features from raw weather data

Tools To Be Used: FeatureTools, Scikit-learn

Approach:

- Create features manually. Potential features include – Public Holiday, Weekend, Avg. temp of last 3 days, etc.
- Use applications like FeatureTools to generate features automatically and apply correlation or Chi sq. test to determine the usability of those features.

FUTURE WORK

MODEL CREATION

Task: Create Statistical and Deep Learning Models with least error and which can work on multiple hierarchical inputs

Tools/Models To Be Used: FB-prophet, H2O.ai (Auto-ML), ARIMA, etc.

Approach:

- Based on the features selected, we would first try and use Auto-ML to see if it comes up with any good models.

- Depending upon the output, we would then proceed with implementing multiple models to find out the best fit for our data.

- Creating dashboard where the user sees the impact of weather on hierarchical features like category-province or department-city, etc.

RISKY BEHAVIOUR

WHAT ARE THE RISKS?

The data being not large enough, it could work for upper level features such as category or department, while there might not be enough data points to train for each style.

HOW DO WE PLAN TO MITIGATE THOSE?

We would create an interactive dashboard where the user can select various combinations to view their predictions. The predictions would be followed by their confidence levels. This can help the user make informed decisions.