DSI x L'Oréal Capstone: Time Series Sales Prediction for Kiehl's

Key words: time series, prediction model, lightGBM

Team: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu (Captain)

Kiehl's SINCE 1851

Columbia University

Motivation

Resource Allocation

Inventory Management

Seasonal Trend Prediction

Business Planning

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Research Questions

30 days sales prediction for Kiehl's stores across the United States

- Can we identify correlation between traffic and sales?
- Can we find determinants and external factors to build traffic model?
- Can we build sales prediction model based on traffic model?

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Data

Data Description		
	External Dataset	Internal Dataset
	Census Data (state & city)	Sales Data (country level)
		Traffic Data (country level)
	U.S. Holiday Dataset	U.S. Traffic Data ★

Correlated

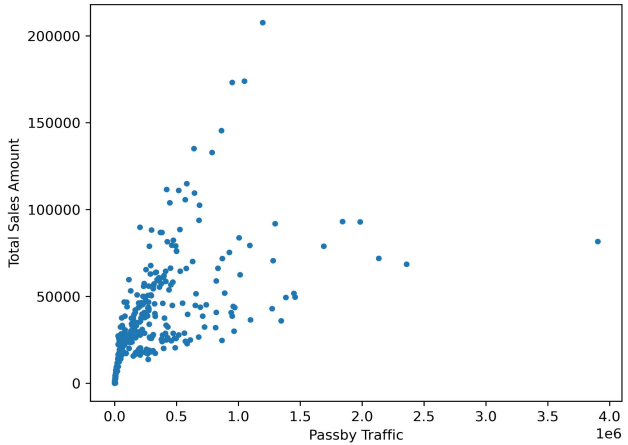


Fig 1: Scatter Plot of Traffic and Sellout

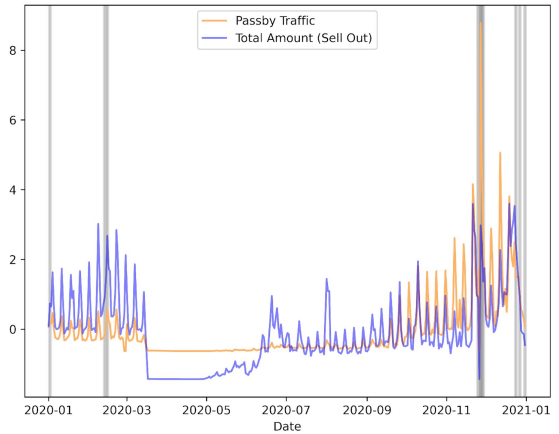


Fig 2: Passby Traffic and Sellout Amount over Time (after standardization)

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Data

- After data cleaning and preprocessing (before encoding), the dataframe involved in this project contains 89317 observations with 17 variables, ranging from 06/01/2016 to 08/18/2019
- **Categorical variables:** state, city, store type, flagship, sales zone, brand, holiday, weekday, year, month
- **Numerical variable:** number of entry (traffic), population (state-level), median household income (aka MHI, state-level), population density (city-level), mhi (city-level), income per capita (city-level)

From external census data source

Day	State	City	Store type	Flagship	Sales Zone	Brand	NbEntry	Holiday	Weekday	Year	Month	Population	MHI	Pop Density	mhi	per capita
2016/1/1	AZ	SCOTTSDALE	Boutique	Standard	Mall	KIEHL'S	40	New Year's Day	Friday	2016	1	62.9	65913	1311.7	97409	70040
2016/1/1	CA	ARCADIA	Boutique	Standard	Mall	KIEHL'S	71	New Year's Day	Friday	2016	1	253.7	84097	5187.2	99588	47167
2016/1/1	CA	BERKELEY	Boutique	Standard	High Street	KIEHL'S	22	New Year's Day	Friday	2016	1	253.7	84097	11917.3	97834	56168
2016/1/1	CA	BREA	Boutique	Standard	Mall	KIEHL'S	52	New Year's Day	Friday	2016	1	253.7	84097	3889.3	108721	47945
2016/1/1	CA	CANOGA PARK	Boutique	Standard	Mall	KIEHL'S	25	New Year's Day	Friday	2016	1	253.7	84097	9942	84535	35589

After transformation, there are 87126 observations with 184 variables

Feature Engineering

Starting Point
Baseline Model

Predicts traffic by taking the average of traffic in historical data on a given date.
RMSE: 25.0

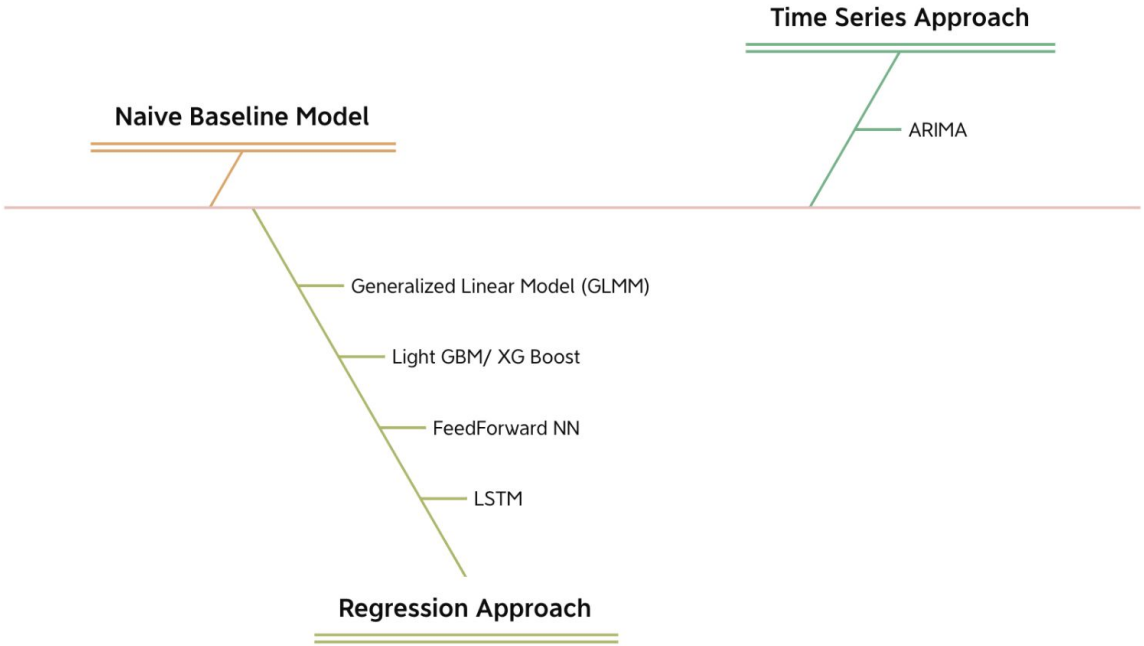
Data Augmentation

- One-hot Encoding
- Geographical Encoding
- Holiday Labels
- Day of the Week

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Methodology

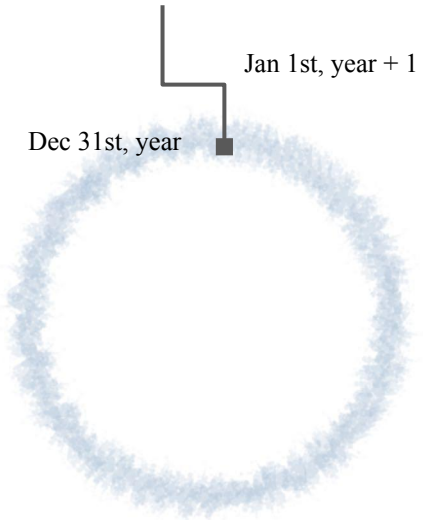
How to construct prediction model?



Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Methodology

Sine and Cosine Transformation



Lookback variables

Date	Nb of entry	1 day lookback	N day lookback
Jan 1st	Y_{t-2}	Y_{t-3}	Y_{t-2-N}
Jan 2nd	Y_{t-1}	Y_{t-2}	Y_{t-1-N}
Jan 3rd	Y_t	Y_{t-1}	Y_{t-N}

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

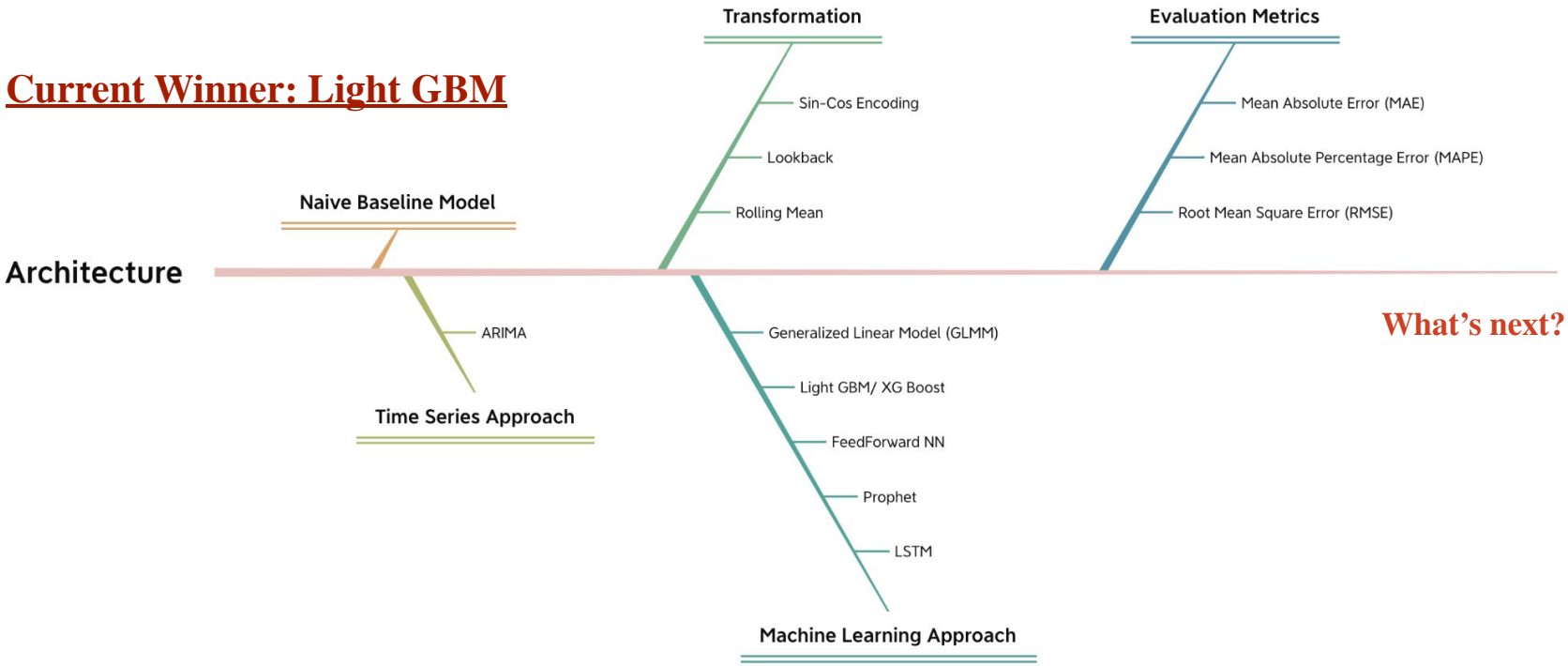
Model Evaluation

Model Comparison	MAPE	MAE	RMSE
Light GBM - Sin-Cos transformation on 365 day - Lookback on 30 days - Rolling window	19.82	5.13	8.19
Light GBM	20.36	5.30	8.43
XGBoost			14.82
FeedForward NN			12.82
Generalized Linear Model (GLMM)			23.93
Prophet (with holiday parameter)			22.09
Baseline (fitting on previous year)			25

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Results

Current Winner: Light GBM



Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Discussion

Is there any limitations besides methodology itself?

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Discussion

How to combine this traffic prediction with sales model?

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Capstone Mentor

Rémi Ferreira

Team Members

Yifan Lu (yl5113@columbia.edu)

Jianing Yu (jy3266@columbia.edu)

Ruijie Zhang (rz2596@columbia.edu)

Yancheng Zhang (xz3157@columbia.edu)

Yifan Zhu (yz4360@columbia.edu) - Team Captain

Thanks for Listening

Columbia University: Yifan Lu, Jianing Yu, Ruijie Zhang, Yancheng Zhang, Yifan Zhu

Reference

[1] Tingyan Deng, Yu Zhao, Shunxian Wang, and Hongjun Yu. Sales Forecasting Based on LightGBM. 2021. https://ieeexplore.ieee.org/document/9342445/references#references

[2] Zhang He, Sun Yu. Application of LightGBM and LSTM combined model in vegetable sales forecast. 2020. https://iopscience.iop.org/article/10.1088/1742-6596/1693/1/012110/pdf

[3] Youyang Zhang, Changfeng Zhu, Qingrong Wang. LightGBM-based model for metro passenger volume forecasting. 2020. https://ietresearch.onlinelibrary.wiley.com/doi/epdf/10.1049/iet-its.2020.0396

[4] Ramos P, Santos N, Rui R. 2015. Performance of state space and ARIMA models for consumer retail sales forecasting[J]. Robotics and Computer-Integrated Manufacturing,2015,34:151-163. https://www.sciencedirect.com/science/article/abs/pii/S0736584515000137

[5] Iván Vallés-Pérez, Emilio Soria-Olivas, Marcelino Martínez-Sober, Antonio J. Serrano-López, Juan Gómez-Sanchís, Fernando Mateo. Approaching sales forecasting using recurrent neural networks and transformers. 2022. https://www.sciencedirect.com/science/article/pii/S0957417422004146