Bakery Sales Prediction: Time Series & Demand Planning

[Team M]

Charbel Imad
Dmitrii Timoshenko
Irina Degtiareva
Zack Brodtman

Berkeley
UNIVERSITY OF CALIFORNIA

Motivation

Goal: Provide a recommendation for bakeries in demand forecasting and planning by

a) **Predicting total sales one week ahead**
b) **Predicting individual daily sales of most popular items**

- Large enterprises implement forecasting models for demand and logistics planning, but it is challenging for small businesses
- We are focusing on popular items that generate most of the revenue to increase operations efficiency
- Demand Forecasting can help to reduce food waste & improve financial and environmental sustainability

Berkeley
UNIVERSITY OF CALIFORNIA

Data

Data:

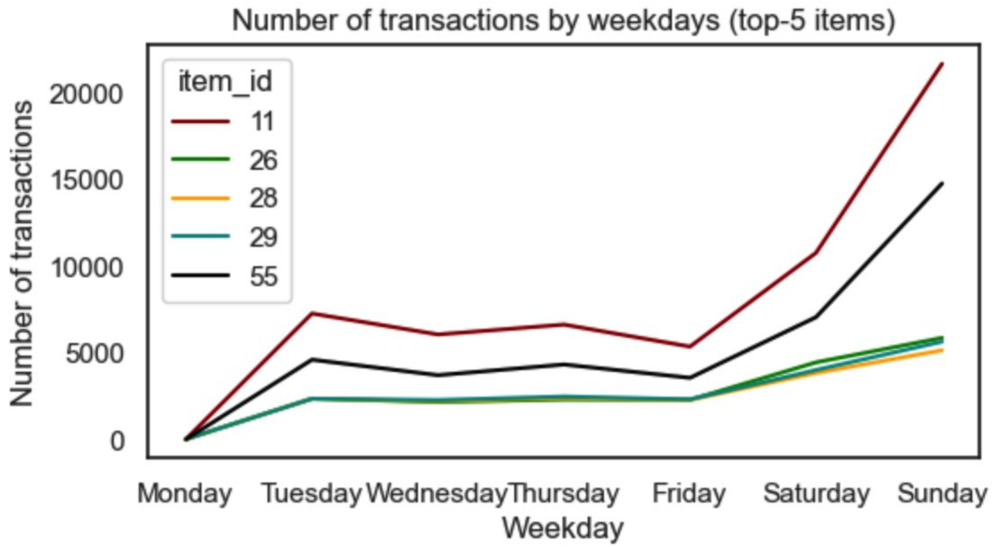
- ➔ 2 Bakeries (South and North France)
- ➔ 4M rows of transactional data & covariates from 2018 to 2022
- ➔ Each observation is a unique purchase with a timestamp
- ➔ Items sharing one transaction_id were purchased together

Covariates	
Variable	**Meaning**
is_holiday	Holiday Flag
holiday_id	Holiday Type
temp	Air temperature
cloud_cov	Cloud coverage

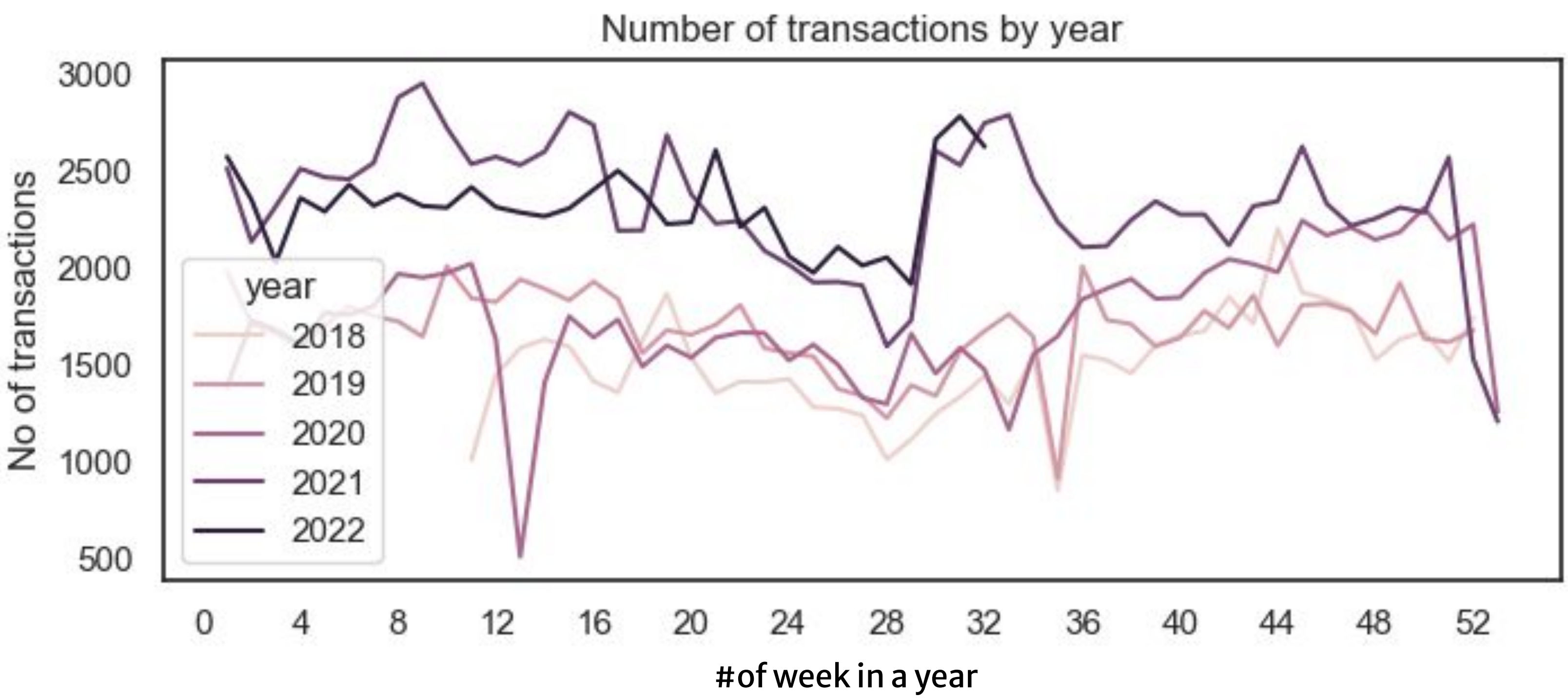
Main Data	
Variable	**Meaning**
item_id	Unique name of item
transaction_id	Unique ID of transaction
quantity	Continuous or discrete number of product sold
price_ht	Total price for a item (including discounts, coupons)
date	Date and time of an observation

EDA – Overview

1. Analyzed the number of transactions by different measures
   a. **COVID** year **didn't affect sales trend**, although there was a gap
   b. Trend has been almost the same for all five years
   c. **Increase** in sales **in the last year**
2. Identified 5 **key items** (out of 100+) for which we can make predictions
3. Trends in sales are the same for the most popular items
   a. Some items are **seasonal**
   b. Decided to exclude not popular items
4. Sales for some items depend on the **weather**



EDA – Historical transaction data



Conclusions: almost the same trends during observed years; gaps in COVID year; overall increase in sales in 2021 and 2022 years; increase in summer sales (week 28-32) in 2021 and 2022

Modeling Summary

	Weekly	Daily
Total Sales	**Statistical Models (OLS + Time Series)**	**Deep Learning**
Individual Items' Sales	**Statistical Models (OLS + Time Series)**	**Deep Learning**

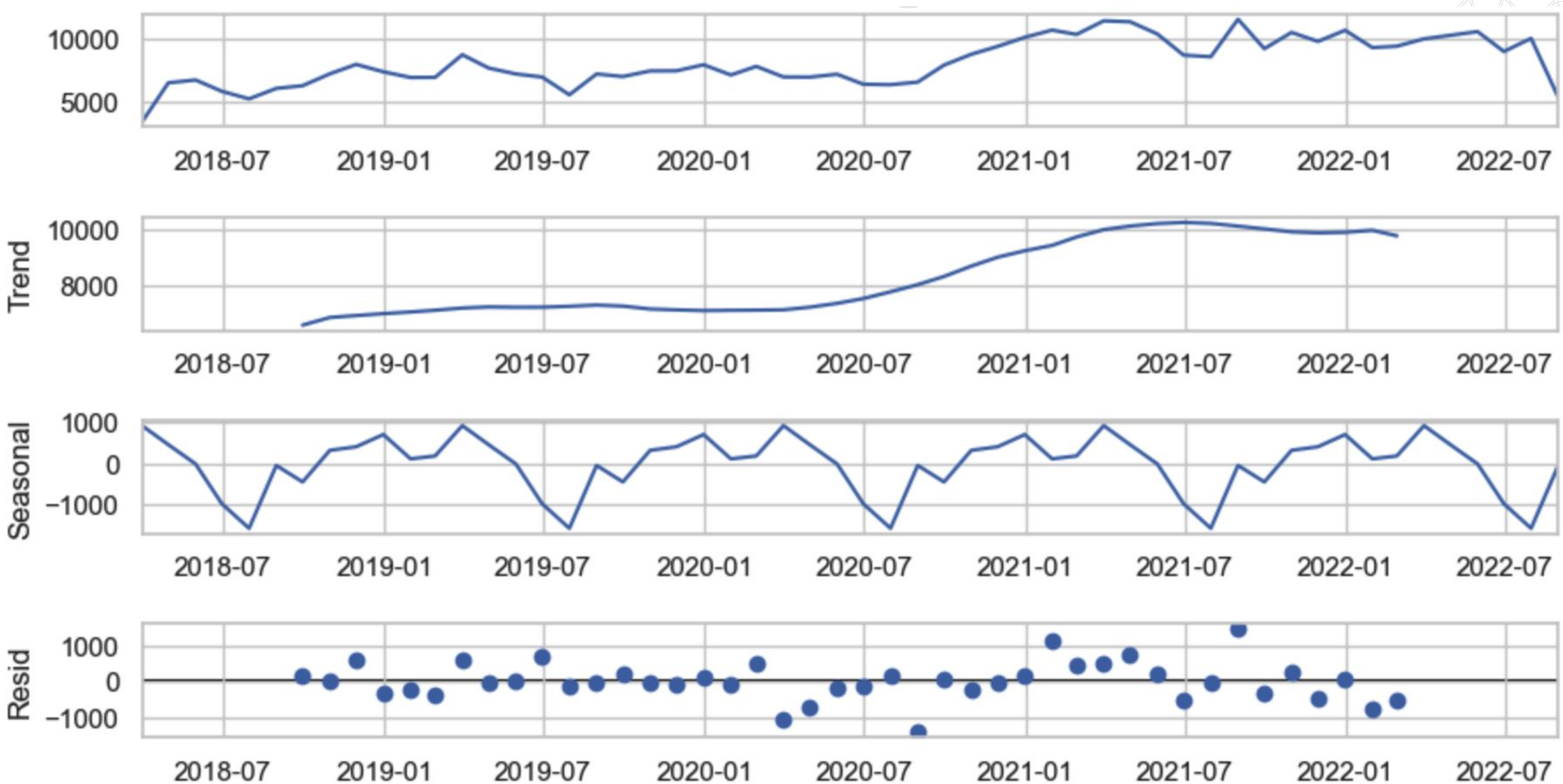
Operational Efficiency & **Revenue** Planning

Demand Forecasting & **Sustainable** Growth

Baseline Prediction Model

Experimental Advanced build-up

Weekly Sales Model



There are **three** components we can use in our analysis:
Trend + **Seasonality** + **Error**

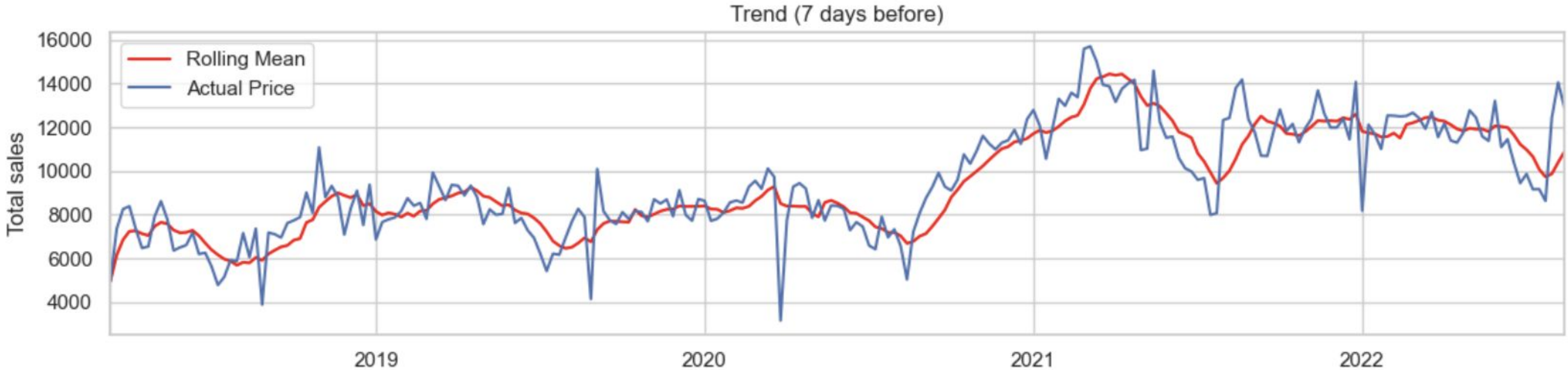
Weekly Sales Model

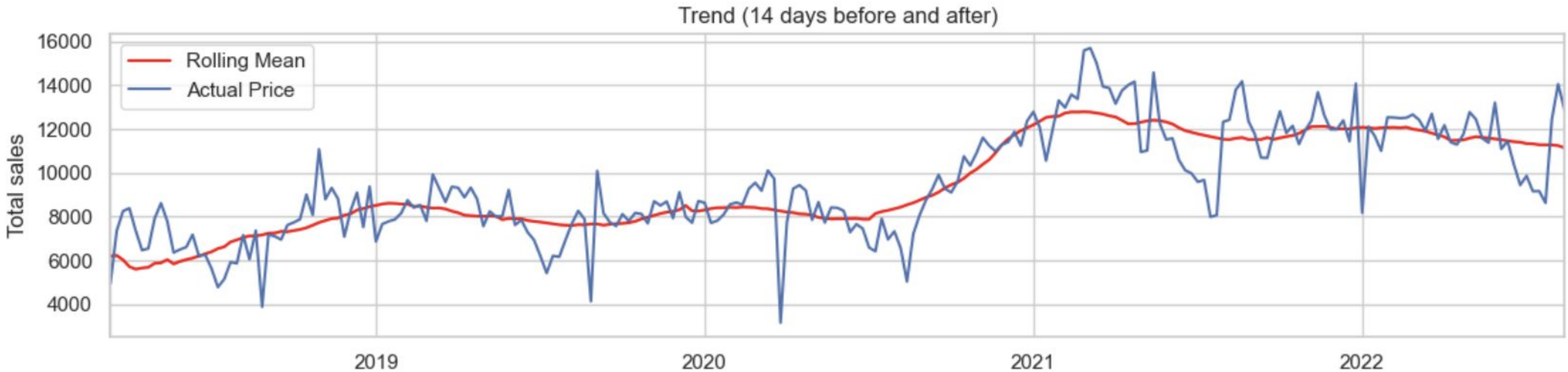
$$Y_t = TR_t + Season_t + Z_t$$

1. Use **rolling mean** to predict trend of data TR_t
   a. Window = [t−14,t+14]
2. Use dummy variables for months to add **seasonal component** $Season_t$ using OLS
3. Add **weather** and **holidays** as dummy variables
4. Predict **residuals** Z_t using autoregressive model

Main challenge is to work with **COVID data** that adds more **uncertainty** to the model prediction

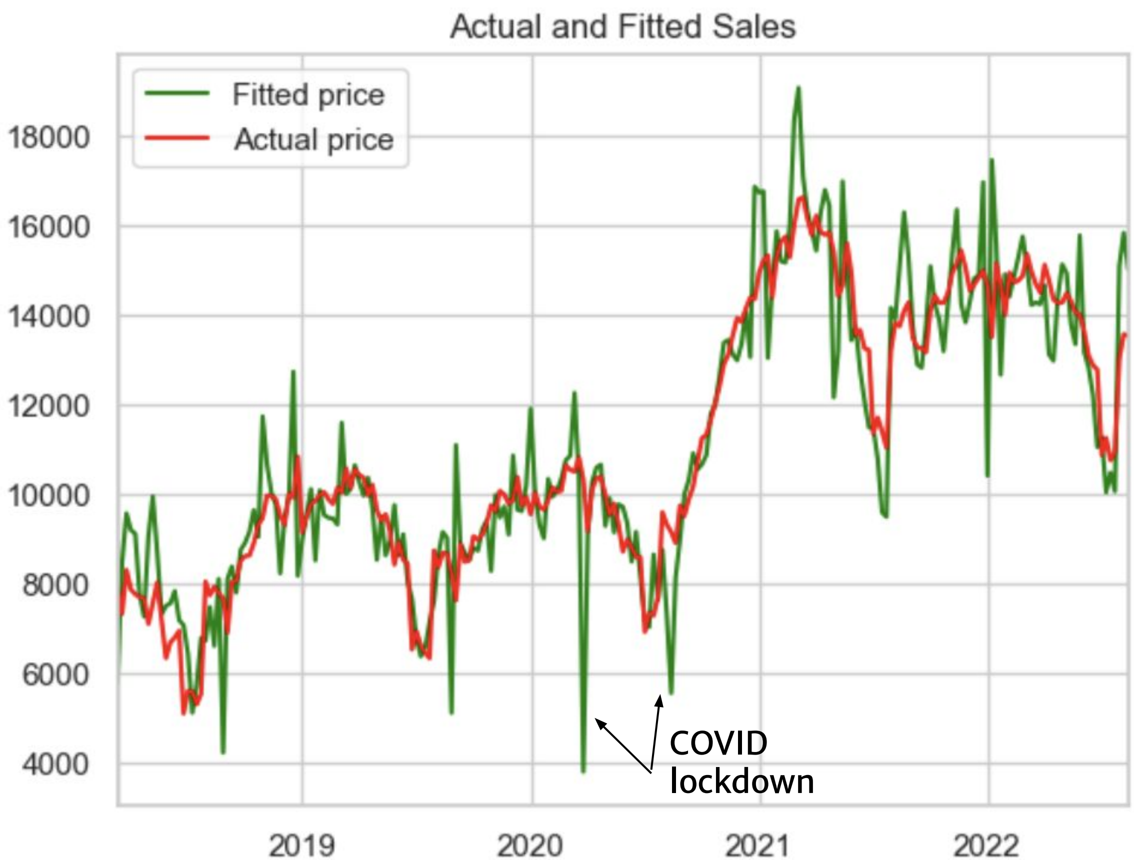
Weekly Sales Model - Trend





- ➔ Using **rolling mean** to extract the trend component
- ➔ Wider window allows to get a **smoother** trend compared to other methods we've tried

Weekly Sales - Seasonality



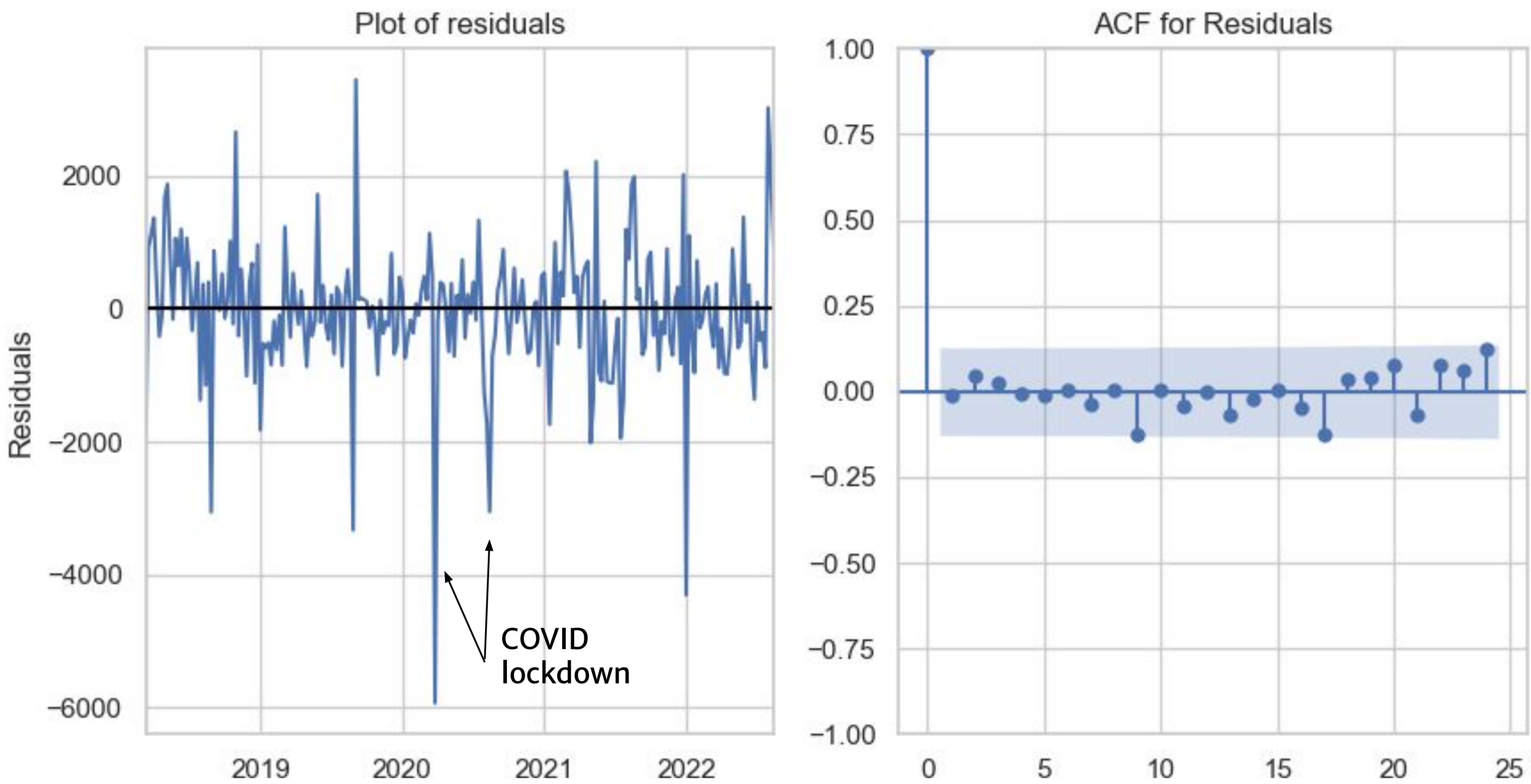
- ➔ There is a **seasonal decrease** in sales every **July**
- ➔ We **can not predict lockdown** spike based on seasonality only
- ➔ Added **holidays** and **weather** to the model

Coefficients from OLS fit

Month	**Coefficient**
Feb	471.7
March	898.1
April	1453.8
May	1025.6
June	203.9
July	**-1409.5**
August	689.6
September	172.9
October	609.3
November	819.1
December	543.2
Holiday	**31.2**
Temp	**-32.5**

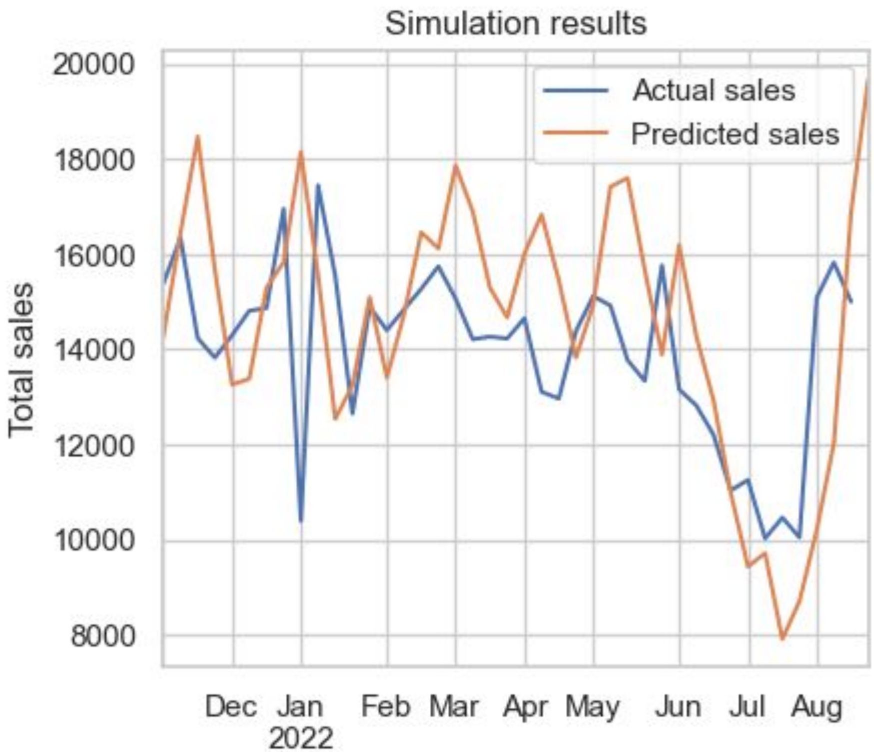
R^2: 0.25

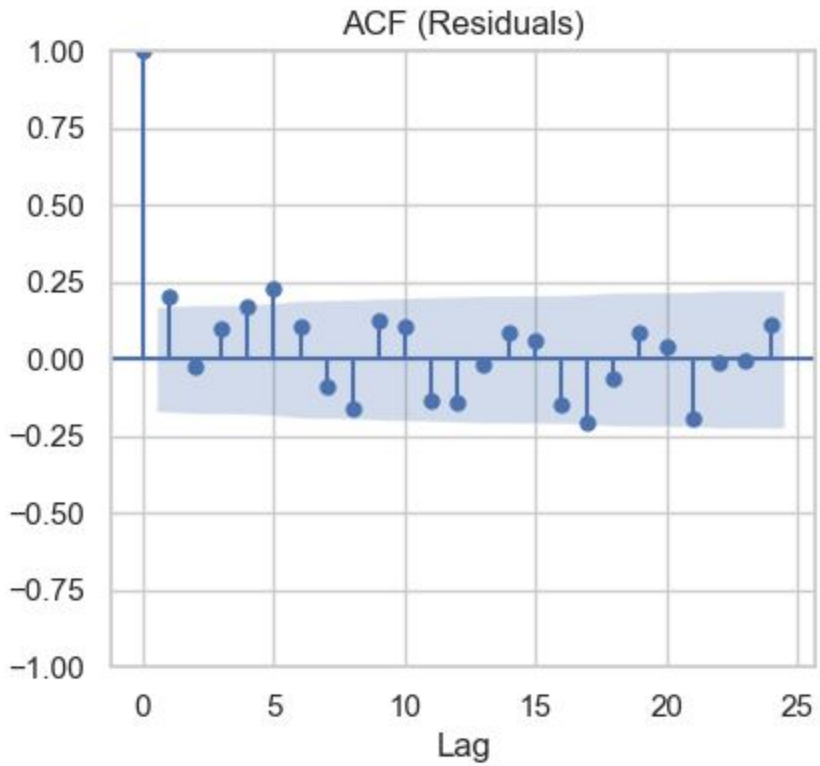
Weekly Sales - Modeled Residuals



- ➔ Noticeable COVID and **unpredicted events**
- ➔ **Residuals** can be modeled as a Time Series **AR(1)** process
- ➔ **Centered** around zero

Weekly Sales - Backtesting*





- ➔ Starting with first 190 weeks **predict the next observation** t+1
- ➔ **Update** trend, model weights and data
- ➔ Starting with first 191 weeks **predict the next observation** t+2
- ➔ …

*MAPE: 13%

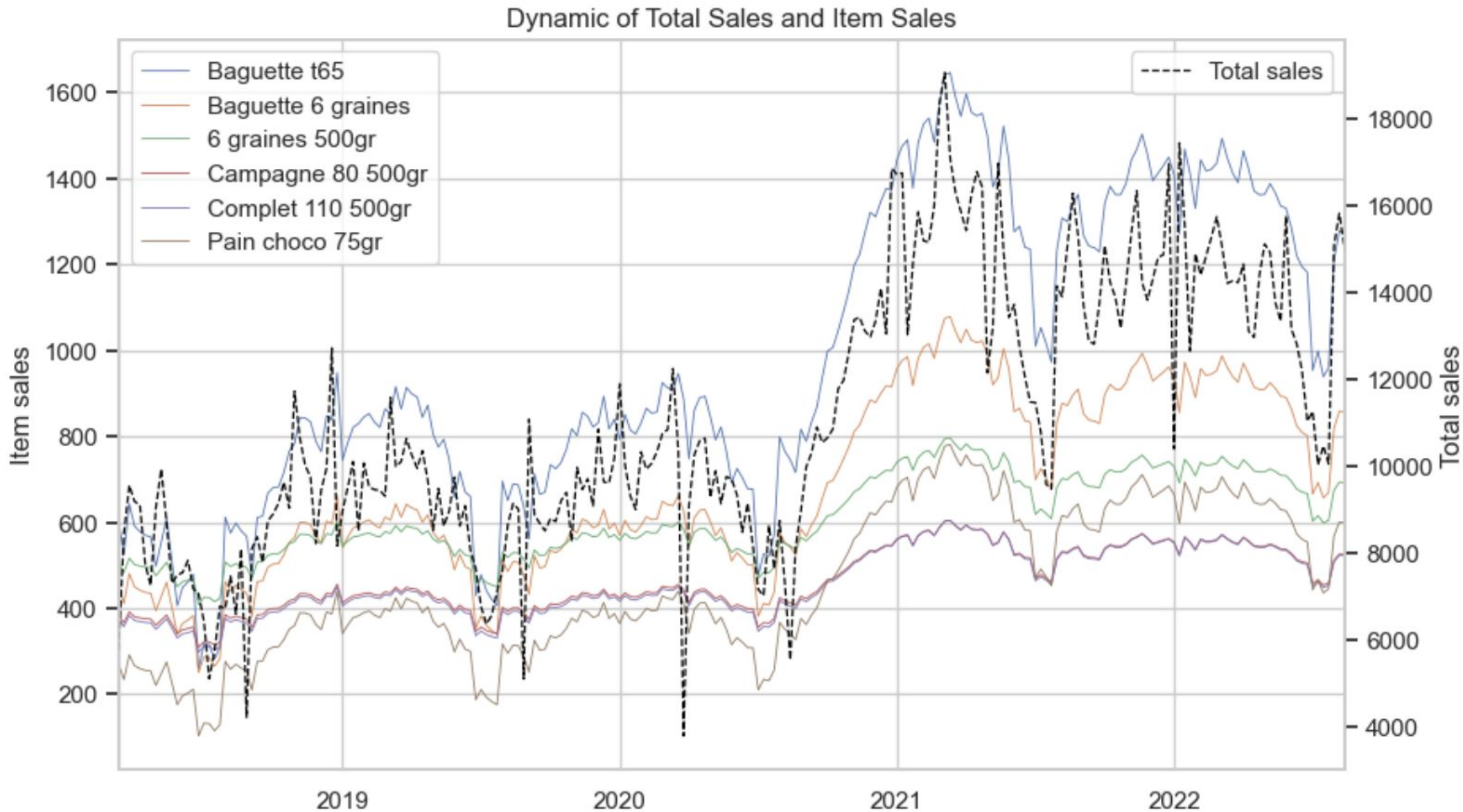
From Total Sales to Individual Items

$$X_{it} = \alpha_0 + \beta Y_t + Z_t$$

Where **X** is individual item's sales and **Y** is a modeled total sales value

- ➔ Predict individual item's sales based on total sales value
- ➔ Using **total sales** predictions we can estimate **individual item's contribution**

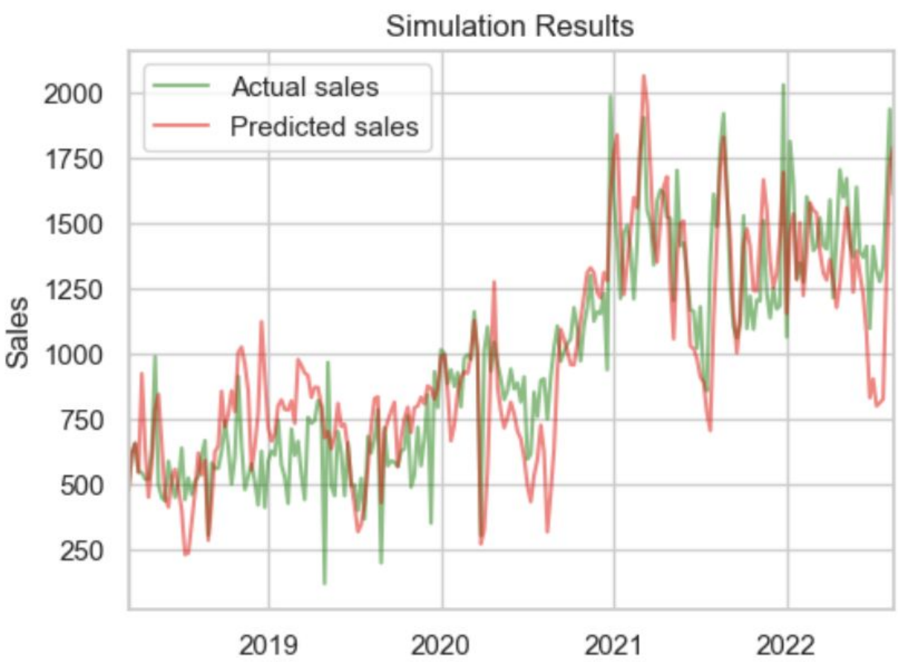
Individual Items*

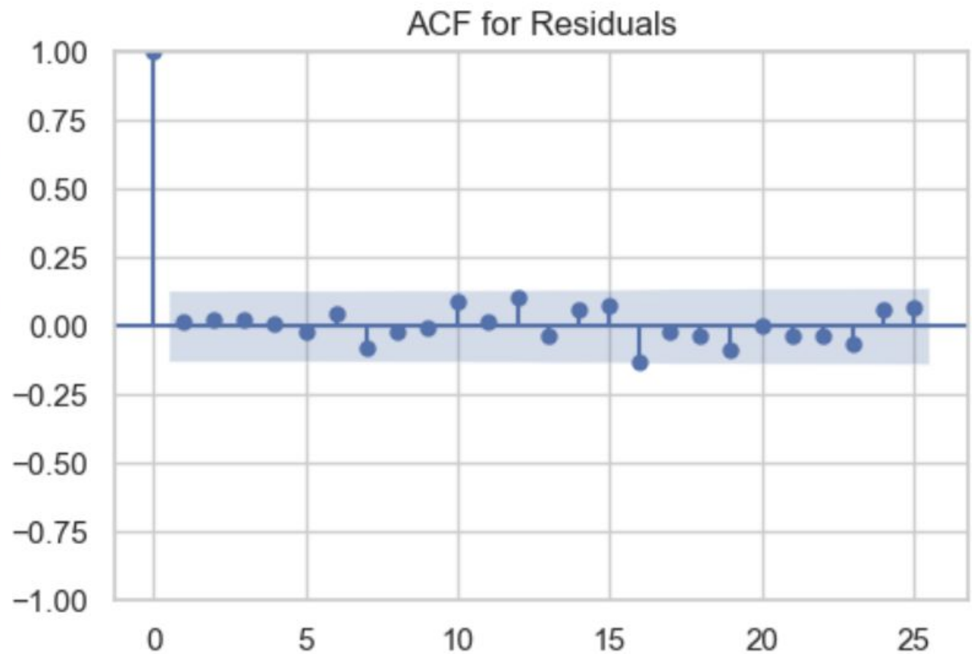


- ➔ Trend of **individual items** is very similar to the trend for **total sales** for the most popular items
- ➔ We can use similar models for total sales and individual items

*all coeffs are significant. R^2 from 0.35 to 0.71

Predicting La Baguette - Weeks





- ➔ Backtesting model for La Baguette starting from 2018
- ➔ Usually sales increase **before** and **after** the holidays (Christmas 2022 spike **was not** predicted)

Daily Total Sales

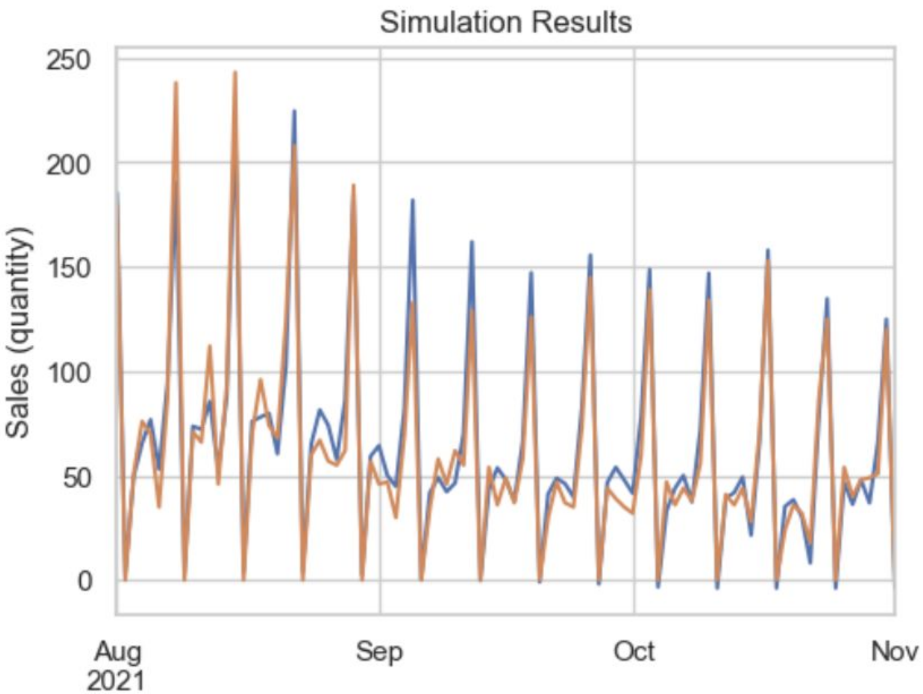
$$Y_t = TR_t + Season_t + Z_t$$

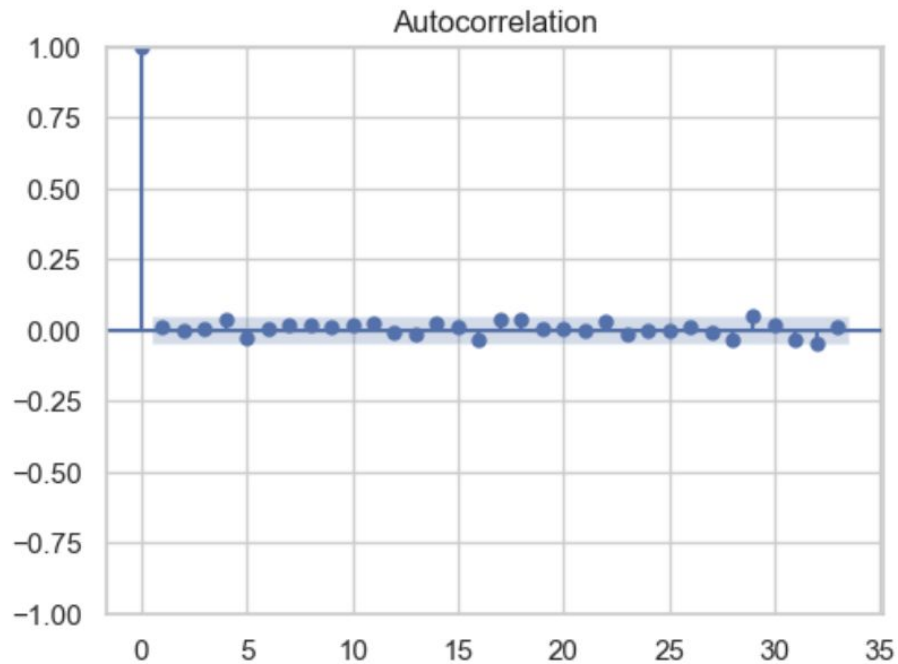
1. Use **rolling mean** to predict trend of data ***TR_t*** *
2. Use dummy variables for days to add **seasonal component $Season_t$** using OLS*
3. Predict **residuals Z_t** using autoregressive seasonal model (SARIMAX)

*R^2: 0.88. Coefficients are signifi
with a=0.01

Berkeley
UNIVERSITY OF CALIFORNIA

Fitting La Baguette - Daily





- ➔ From EDA we know that bakery is **closed on Monday** and peak sales are on weekends
- ➔ Same pattern for every week (except holidays)
- ➔ Now **seasonal** covariates are **days of the week**

Limitations

- External factors – our model is trained to ignore the dip due to COVID-19, we can not predict some events like festivals and concerts in the area
- Outliers – we are able to accurately predict individual daily sales of the most popular items, but for less frequently purchased items our model is not appropriate
    - E.g. we are not able to accurately predict Christmas cakes

Need a more advanced model to improve predictions

Solution: Neural Networks

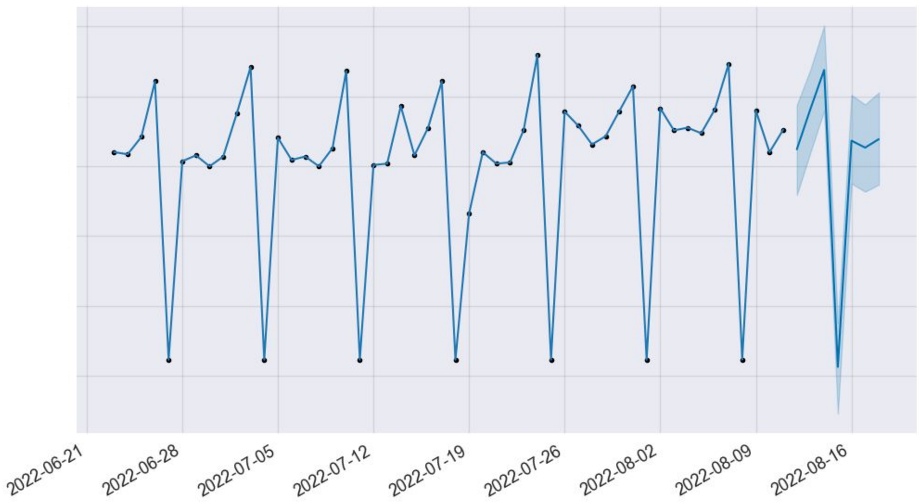
Berkeley
UNIVERSITY OF CALIFORNIA

First model: Prophet

The model

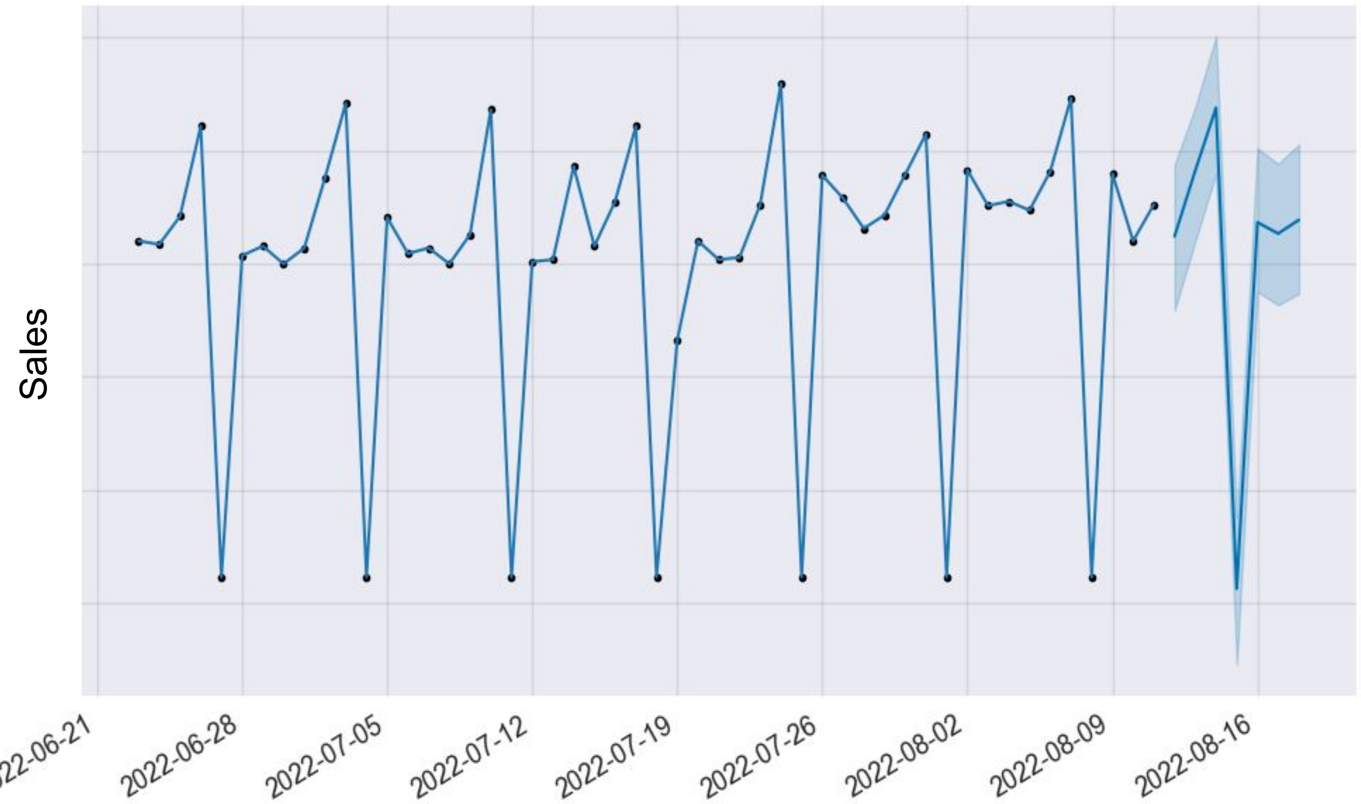
$$y_t = g(t) + s(t) + h(t) + \varepsilon_t,$$

g(t): piecewise-linear trend
s(t): seasonal patterns
h(t): holiday effects



7 day forecast for item ‘Baguette’ of shop #1

First model: Prophet



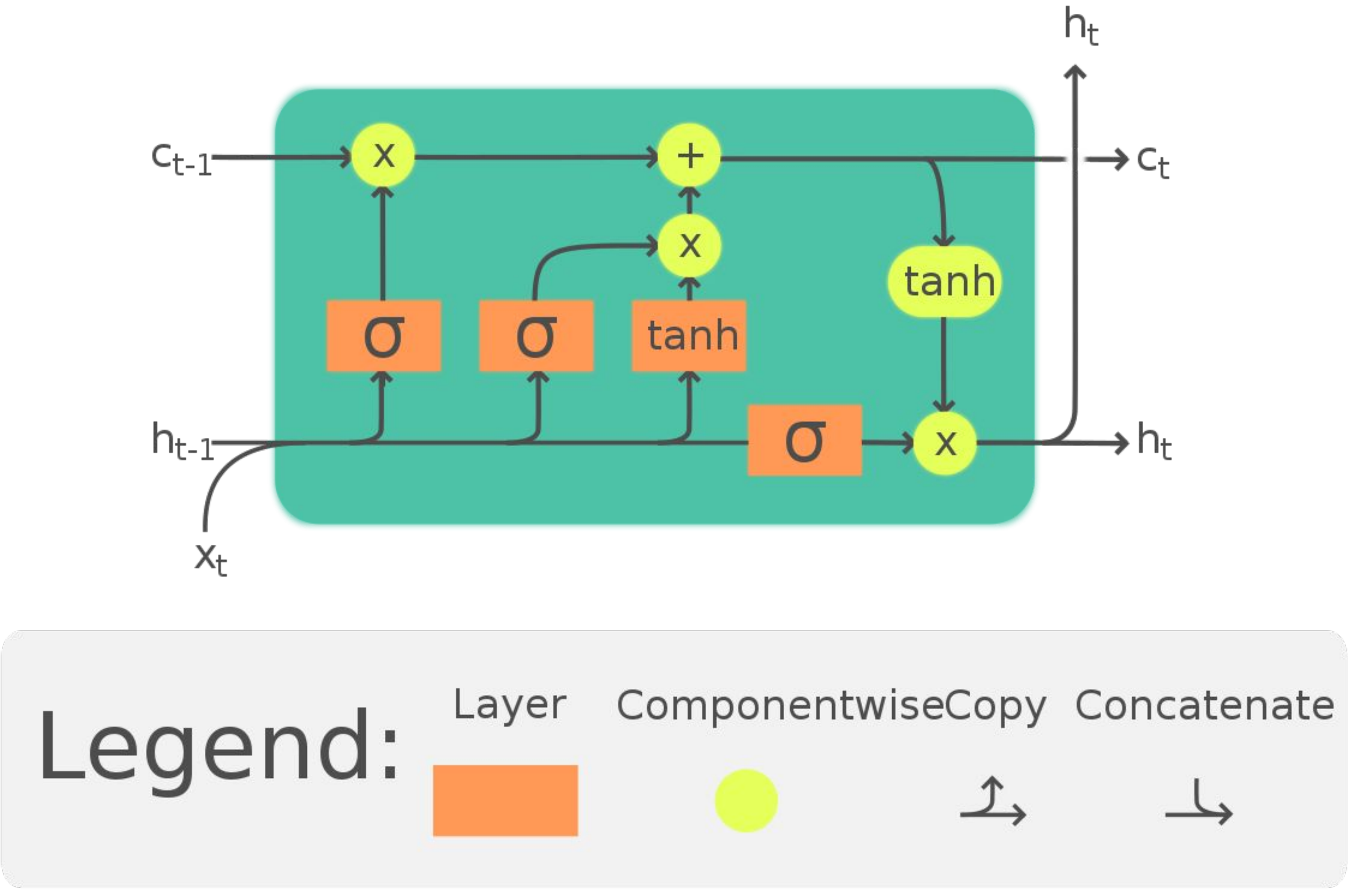
Berkeley
UNIVERSITY OF CALIFORNIA

Second model: Long Short Term Memory (LSTM)

- LSTM is a type of recurrent neural network designed to handle the vanishing gradient problem in sequential data.
- Use a specialized architecture that includes memory cells, input gates, forget gates, and output gates to selectively retain or discard information over time.
- Used in time series prediction to model the underlying patterns and dependencies in sequential data.
- Can make predictions on new data by feeding it a sequence of past observations and generating a forecast for the next value in the sequence.

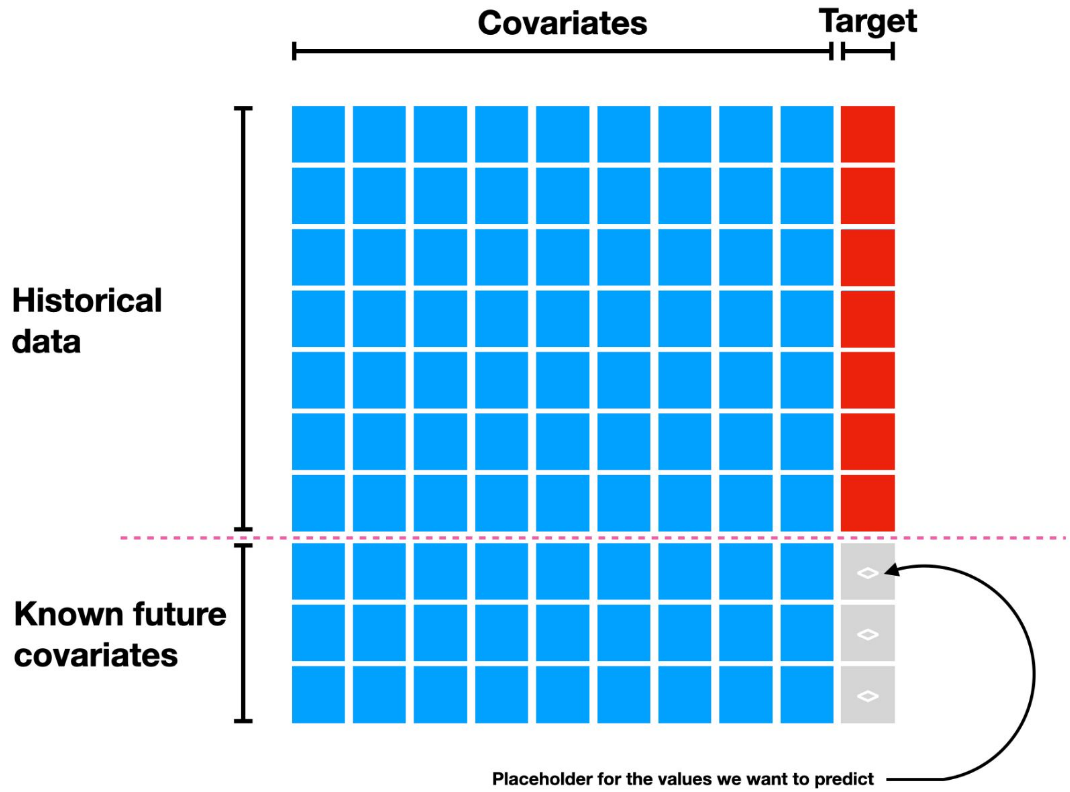
Berkeley
UNIVERSITY OF CALIFORNIA

Long Short Term Memory (LSTM)



Guillaume Chevalier from https://commons.wikimedia.org/wiki/File:The_LSTM_Cell.svg

Using LSTM to forecast daily individual item demand



Formatting the data

Using LSTM to forecast daily individual item demand

We trained the model on one of the two shops, filtered on the top 10 selling items.

	weekday	day_of_month	month	year	…	quantity
2018-03-08	-0.0015	-0.87	-1.02	-1.36	…	1.47
…	…	…	…	…	…	…
2022-07-29	0.499	1.51	0.15	1.64	…	-0.94

The dataframe of input covariates (date information, calendar and weather data) along with the target column (quantity).

Data has been normalized with standard scaling for both the input and target.

Berkeley
UNIVERSITY OF CALIFORNIA

Using LSTM to forecast daily individual item demand

We trained the model on one of the two shops, filtered on the top 10 selling items.

	weekday	day_of_week	day_of_month	month	year	calendar_embedding	temp_min	temp_max	rain	wind	cloud_cov	quantity
2018-03-08	-0.000150	-0.000150	-0.876583	-1.023898	-1.361219	-0.158539	-1.082799	-0.833544	0.529138	2.085198	-0.407358	1.470600
2018-03-09	0.499906	0.499906	-0.762724	-1.023898	-1.361219	-0.158539	-1.082799	-0.833544	0.029604	-0.569900	1.099230	-1.773285
2018-03-10	0.999963	0.999963	-0.648865	-1.023898	-1.361219	-0.158539	-0.168775	-0.423348	0.386414	0.567999	0.784811	-1.773285
2018-03-11	1.500019	1.500019	-0.535006	-1.023898	-1.361219	-0.158539	0.196834	-0.013153	3.407407	-0.064167	0.326285	-1.773285
2018-03-19	-1.500318	-1.500318	0.375864	-1.023898	-1.361219	-0.158539	-2.362431	-1.790666	-0.446143	0.188699	0.186543	-1.773285
...	...	...	...	...	...	...	...	...	...	...	...	...
2022-07-25	-1.500318	-1.500318	1.059017	0.162280	1.641031	-0.158539	1.659271	1.354166	-0.469930	2.085198	-0.071105	-1.773285
2022-07-26	-1.000262	-1.000262	1.172876	0.162280	1.641031	-0.158539	1.110857	0.807238	-0.446143	0.188699	0.304450	-1.773285
2022-07-27	-0.500206	-0.500206	1.286735	0.162280	1.641031	-0.158539	0.196834	1.217434	-0.469930	-0.569900	-1.267641	-0.942658
2022-07-28	-0.000150	-0.000150	1.400594	0.162280	1.641031	-0.158539	0.562443	1.490898	-0.469930	-0.064167	-0.547099	-1.773285
2022-07-29	0.499906	0.499906	1.514453	0.162280	1.641031	-0.158539	1.293662	1.764361	-0.469930	-1.075633	-0.101673	-0.942658

The dataframe of input covariates (date information, calendar and weather data) along with the target column (quantity). Data has been normalized with standard scaling for both the input and target.

Using LSTM to forecast daily individual item demand

LSTM (19712 params)

LSTM (12416 params)

LSTM (528 params)

LSTM (119 params)

Parameters in total: 32,775

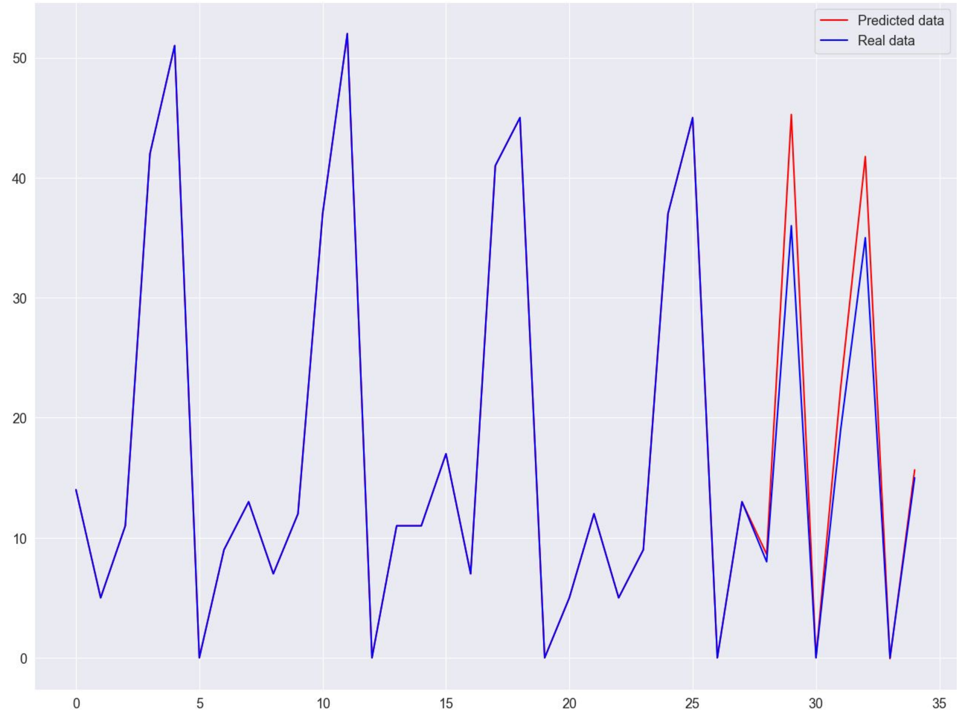
Training dataset: 10k samples
Epoches: 50
Batch size: 64

Using LSTM to forecast daily less popular item demand

Forecasted data on the test set from shop # 1.
MAPE: 29%, MAE: 4.5

Berkeley
UNIVERSITY OF CALIFORNIA

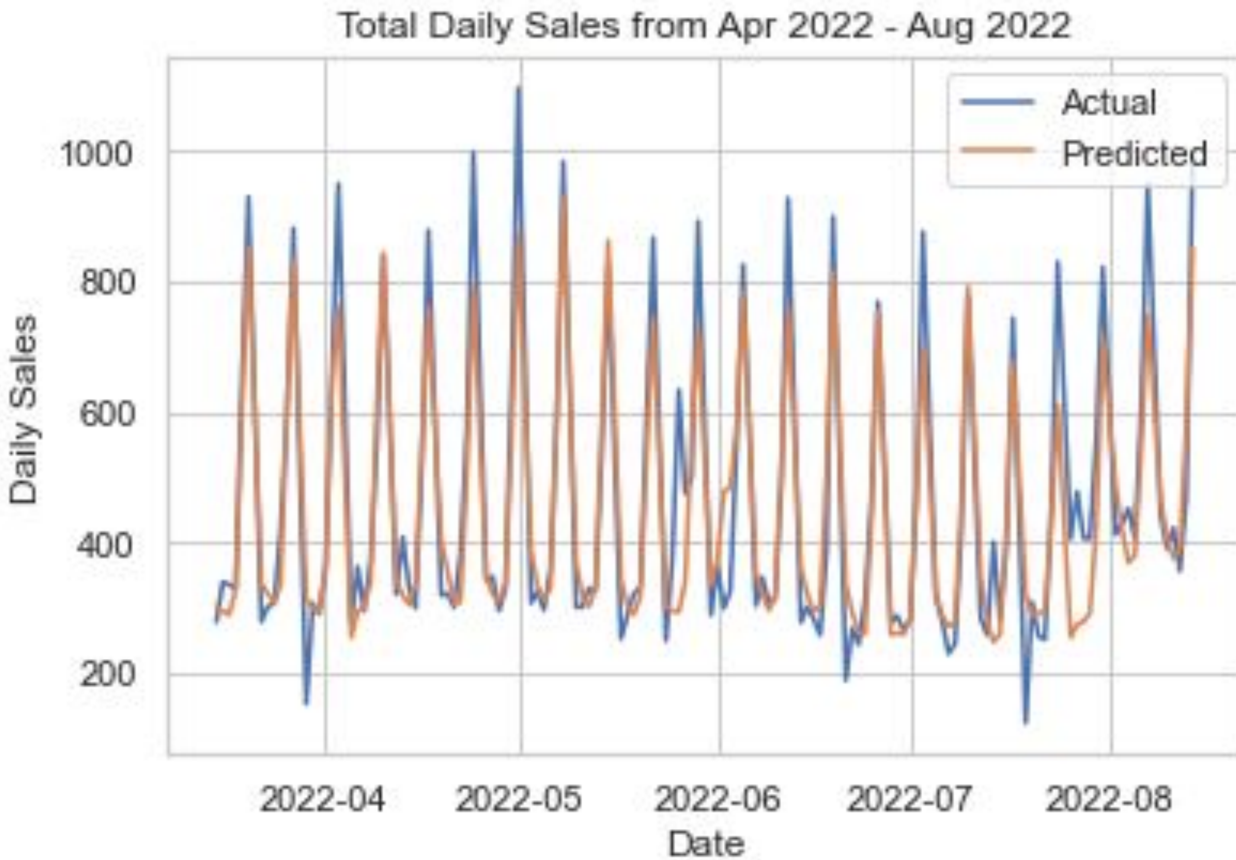
Using LSTM to forecast daily less popular item demand



Forecasted data from shop #2.

Evaluation results on all of the data of the shop: MAPE: 28%, MAE: 4.3

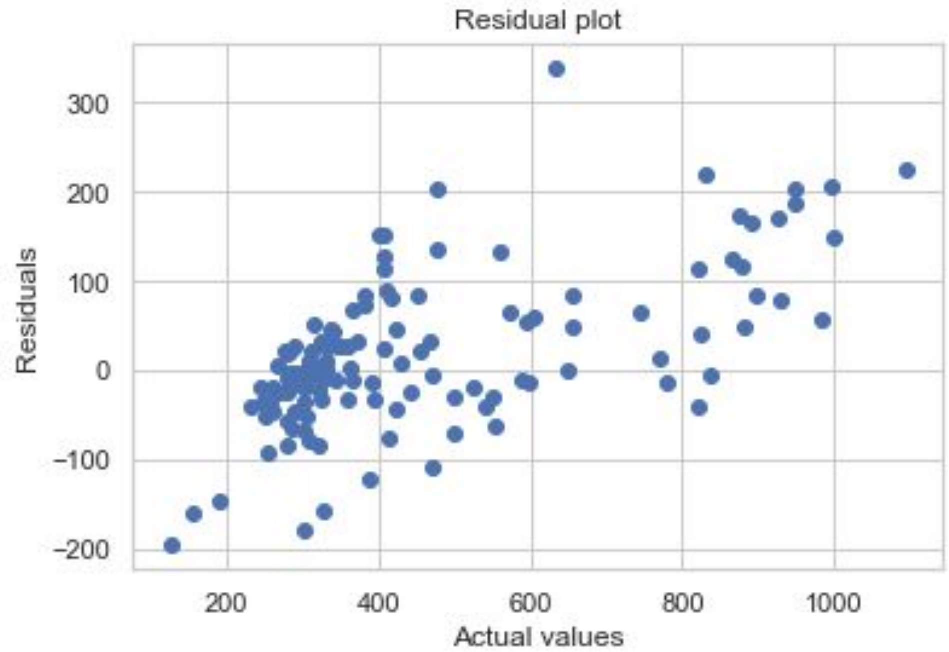
Long Short Term Memory (LSTM)



RMSE: 87.03
MAPE : 14.87%

Berkeley
UNIVERSITY OF CALIFORNIA

Long Short Term Memory (LSTM)



Clear heteroskedasticity – but centered around 0

- extreme events harder to predict
- model overpredicts smaller values and underpredicts larger values

Less Successful Models

ARIMA - MA(1), AR(2), mentioned in last presentation

- limited capacity to model complex nonlinear relationships in the data, such as those with long-term dependencies and nonlinear trends
- They are often only suitable for modeling linear relationships and stationary time series
- Not appropriate for daily sales data that can exhibit seasonal, trend, and irregular patterns.

Different trends

- Worse predictive power
- Less smooth

Conclusion

- Survey data suggests employees have a forecasting accuracy of 60–70%
- Our weekly models have accuracy ~87%
- Our daily models have accuracy ~ 85%
- Best performing models can reach 99% – but these require multiple days and $$$ to train
- Statistical models highly interpretable and generalizable however:
  - LSTM models have better accuracy for daily sales predictions
  - Doesn't give multivariate output – need to create a new model for each item or shop
  - LSTM are better for scaling

Berkeley
UNIVERSITY OF CALIFORNIA