Hot Dog!

Rice Datathon 2021

Kai Hung, Luke Stancil, Brian Xu

Understanding the Data

Looking at the Meaning of Chevron's Hot Dog Sales

- A typical hot dog costs $.25 to make[1]
- Although the margins are not affected substantially by making too many hot dogs, assuming just 10 hot dogs are thrown out each day, with 7800 Chevron® and Texaco® locations in the United States, we can estimate Chevron® loses $28.5 million dollars a year due to unsold hot dogs.
- However, with a profit margin of around $1.75, it is still best to estimate on the high end for hot dog quantities to make.

[1]Webstaurant Store

Assumptions

Assumptions That Formed Our Methodology

- The typical hot dog purchase is not one of impulse, but of routine
    - Given the dataset, we must adhere to a simpler model in order to predict the number sold per hour bucket at the respective stores. After examining different data analysis tools, using a simple median was the most accurate measure.
- Trends are different for different sections of the week, but trends repeat throughout the entire year.
- Because we are given two random days as our testing data, we must attempt to identify which section of the week they fall in and predict accordingly.

Data to Include For A Nationwide Plan

Data for a Brand-wide Hot Dog Cooking Plan

- Distance From Highways
- Gas Station Saturation
- City Demographics
- Chevron’s Market Share in that Location

Methodology

- First, we plotted the data to make sure that our assumptions were correct.
  - For each period of 7 days on the graph there was a 3-day spike and a 4-day drop, which confirmed that hot dog sales vary considerably depending on the part of the week.
- Because the aim of this project was to predict hot dog sales for each respective store for each of the 4 time bucket, we first grouped by store, then found the average number of hot dogs sold for the time bucket.
- Because the information did not tell us when the days were, we accounted for three variables: whether it was in that three day surge, the store, and the time bucket.
  - We used these variables to create a dictionary with the median.
  - Because a true median without the weekday distinction would be highly inaccurate, we looked at the scoring data and found that it was more likely to be a weekday and treated it as such.

Findings

Charts

RMSE = 7.07

Actual

StoreNumber	dayOfTheYear	3HourBucket	GrossSoldQuantity
1000	364	1	0
1000	364	2	3
1000	364	3	4
1000	364	4	1
1000	365	1	9
1000	365	2	9
1000	365	3	3
1000	365	4	6
2000	364	1	1
2000	364	2	2
2000	364	3	1
2000	364	4	2
2000	365	1	1
2000	365	2	1
2000	365	3	7
2000	365	4	3
3000	364	1	0
3000	364	2	1
3000	364	3	1
3000	364	4	2
3000	365	1	1
3000	365	2	2
3000	365	3	4
3000	365	4	7
4000	364	1	6
4000	364	2	11
4000	364	3	30
4000	364	4	24
4000	365	1	12
4000	365	2	16
4000	365	3	18
4000	365	4	37

Predicted

StoreNumber	dayOfTheYear	3HourBucket	GrossSoldQuantity
1000	364	1	5
1000	364	2	5
1000	364	3	5
1000	364	4	5
1000	365	1	5
1000	365	2	5
1000	365	3	5
1000	365	4	5
2000	364	1	5
2000	364	2	5
2000	364	3	5
2000	364	4	5
2000	365	1	5
2000	365	2	5
2000	365	3	5
2000	365	4	5
3000	364	1	5
3000	364	2	5
3000	364	3	5
3000	364	4	5
3000	365	1	5
3000	365	2	5
3000	365	3	5
3000	365	4	5
4000	364	1	15
4000	364	2	27
4000	364	3	36
4000	364	4	45
4000	365	1	15
4000	365	2	27
4000	365	3	36
4000	365	4	45