TE360

Jack Chen

Zihengc2

Integrative Project - 2nd milestone

Part 1: Understanding long-term trends and seasonality & Forecasting with Prophet (30 points)

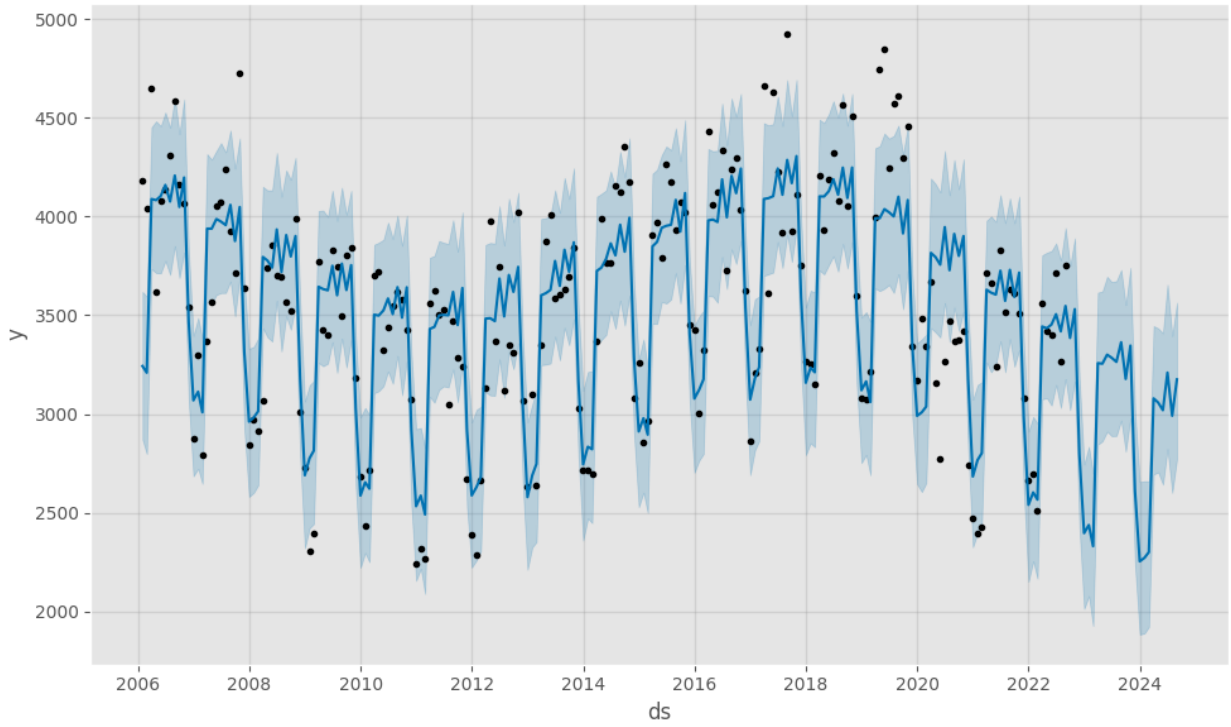
Q 1.1: What are your observations about long-term trends and seasonality? (15 points)

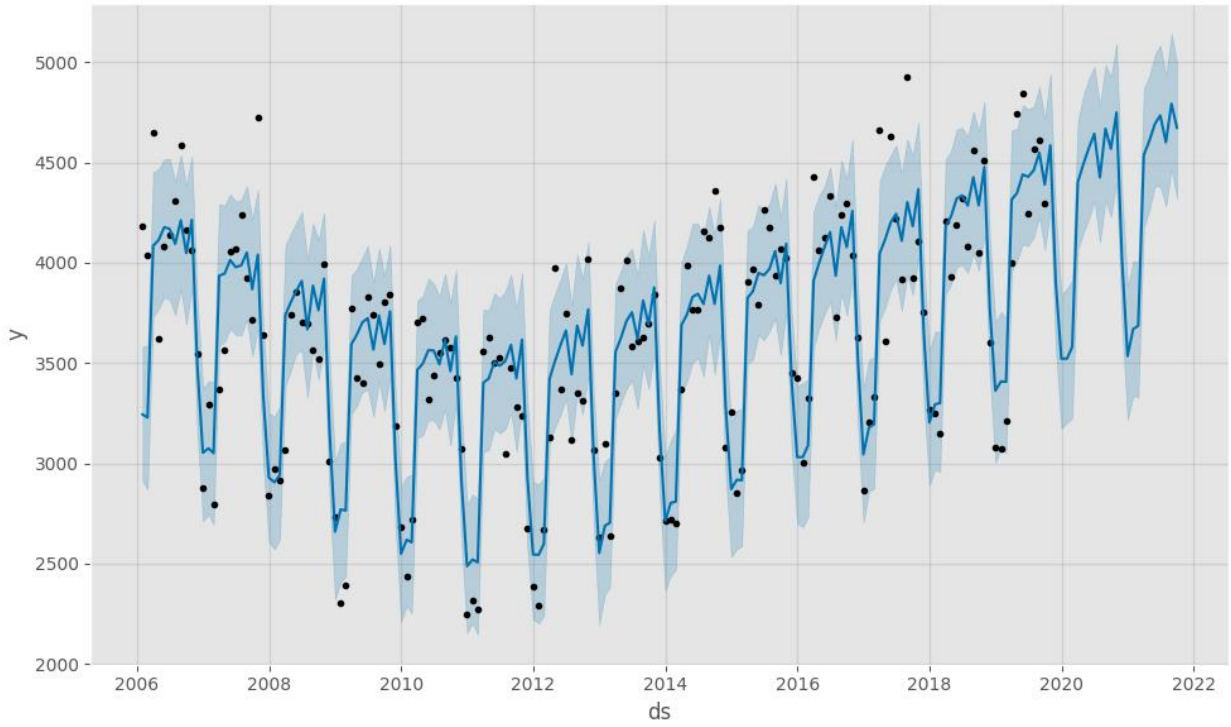
The general trend behaves similar to a sinusoidal wave. It makes sense intuitively since it directly reflects the economics and the demand of commercial space. It slowly dropped when COVID started, and the prediction shows that it will rebound in the near future.

The number of building permits always drops significantly in winter and rebound throughout the year. It makes sense for construction work to slow down during holiday hours in addition to severe cold.

Q1.2: Try forecasting with Prophet using different time periods (different number of months). What do you find interesting about the predictions? (15 points)

If I don't feed the model the data from 2020, the model does not know about the decrease in number of building permits, so it will predict the overall trend to be increasing. However, it still does have the seasonal pattern.


y
ds
5000
4500
4000
3500
3000
2500
2000
2006
2008
2010
2012
2014
2016
2018
2020
2022
2024


y
ds
5000
4500
4000
3500
3000
2500
2000
2006
2008
2010
2012
2014
2016
2018
2020
2022

Part 2: Follow a similar procedure and fill in the TODOs in the code boxes. Your result should be the same as (or similar to) the results shown below the code boxes. (60 points)

TODO 1:

1. Read the data from CSV and save the data frame named "towed_vehicles_df" (5 points)

2. Print the length of the towed_vehicles_df, and print the first 10 rows of towed_vehicles_df(10 points)

	PERMIT NUMBER	ACCOUNT NUMBER	SITE NUMBER	LEGAL NAME	DOING BUSINESS AS NAME	ISSUED DATE	EXPIRATION DATE	PAYMENT DATE	ADDRESS	ADDRESS NUMBER START	...	STREET	STREET TYPE	CI
0	1556602	328992	1	THE LIFEWAY KEFIR SHOP LLC	LIFEWAY KEFIR SHOP	07/16/2021	02/28/2022	07/16/2021	0 W DIVISION ST	0	...	DIVISION	ST	CHICAC
1	1531303	399498	1	JERRY'S SANDWICHES LS, LLC	JERRY'S SANDWICHES	07/16/2021	02/28/2022	07/16/2021	4739 N LINCOLN AVE	4739	...	LINCOLN	AVE	CHICAC
2	1553078	463188	1	ETTA RIVER NORTH, LLC	ETTA	07/16/2021	02/28/2022	07/16/2021	0 N CLARK ST	0	...	CLARK	ST	CHICAC
3	1534556	252742	1	SQUARE KITCHEN, LLC	FORK	07/16/2021	02/28/2022	07/16/2021	4600 N LINCOLN AVE	4600	...	LINCOLN	AVE	CHICAC
4	1556006	337178	1	ROCCO'S, LLC	RANALLI'S	07/16/2021	02/28/2022	07/16/2021	0 N LINCOLN AVE	0	...	LINCOLN	AVE	CHICAC
5	1536621	414414	1	BBSC #4 LLC	BROWN BAG SEAFOOD CO.	07/16/2021	02/28/2022	07/16/2021	3400 N LINCOLN AVE	3400	...	LINCOLN	AVE	CHICAC
6	1559950	34063	1	GASTHAUS ZUM LOEWEN, INC.	THE REVELER	07/16/2021	02/28/2022	07/16/2021	0 W ROSCOE ST	0	...	ROSCOE	ST	CHICAC
7	1540360	23957	1	TEMPO CAFE LIMITED	TEMPO CAFE	07/16/2021	02/28/2022	07/15/2021	6 E CHESTNUT ST	6	...	CHESTNUT	ST	CHICAC
8	1543349	425540	1	MI FOGATA INC.	MI FOGATA INC.	07/16/2021	02/28/2022	07/16/2021	4322 N WESTERN AVE	4322	...	WESTERN	AVE	CHICAC
9	1555187	340126	1	SHINE RESTAURANT CORP.	SHINE RESTAURANT, RISE SUSHI RESTAURANT	07/17/2021	02/28/2022	07/14/2021	0 W WEBSTER AVE	0	...	WEBSTER	AVE	CHICAC

10 rows × 22 columns

TODO 2:

Print the 10 dates ("DATE ISSUED") with the most number of business licenses issued, with their number of licenses issued (15 points)

ISSUED DATE	count
2022-03-31	492
2017-04-30	428
2015-04-30	413
2021-06-30	408
2019-04-30	395
2018-03-31	395
2018-04-30	394
2013-04-30	390
2016-03-31	384
2010-03-31	356

TODO 3:

1. Drop the rows with not-null values(10 points)

2. Extract only the dates ("DATE ISSUED") from towed_vehicles_df, reset the index (15 points)

3. Print the first 10 rows of the result towed_vehicles_df(5 points)

	ISSUED DATE
0	2021-07-16
1	2021-07-16
2	2021-07-16
3	2021-07-16
4	2021-07-16
5	2021-07-16
6	2021-07-16
7	2021-07-16
8	2021-07-16
9	2021-07-17

Part 3: Time series & Simple linear regression (90 points)

Q3: Turn in an example where you compare two data sets using regression analysis. It can be a simple linear regression as Professor Singer did in the class or more sophisticated if you would like. You can use Python or Matlab or even Excel if you want.