

HOTEL TIME SERIES MODELING

BY: *Paranut Prasittipap*





CONTENT

01

Problem Capturing

02

Data Collection

03

Data Preprocessing

04

Exploratory Data
Analysis (EDA)

05

Modeling

- Model Evaluation
- Model Interpretation
- Model Forecasting

06

Result Analysis

01

Problem
Capturing

BACKGROUND



TUBTIM SIAM SUVARNABHUMI HOTEL

A **business hotel**
located near Suvarnabhumi airport



01

Problem
Capturing

PROJECT IDEAS

Problems / Ideas

- The hotel industry has been severely impacted by the Covid-19 lock down situation, resulting in **how much revenue loss ?**
- Is the hotel's total revenue during 2023 on an **increasing trend ?** If so, will the total revenue be **higher than** before Covid-19 situation occurred ?
- What **strategies** should the hotel focus to increase total revenue? (Ex: increasing the room price/pushing marketing)
- Is it still **worth investing** in Online Travel Agencies?

Assumption

- Covid Situation will be assumed to be 1 year, during **2020-03-29** to **2021-03-28**

Goal

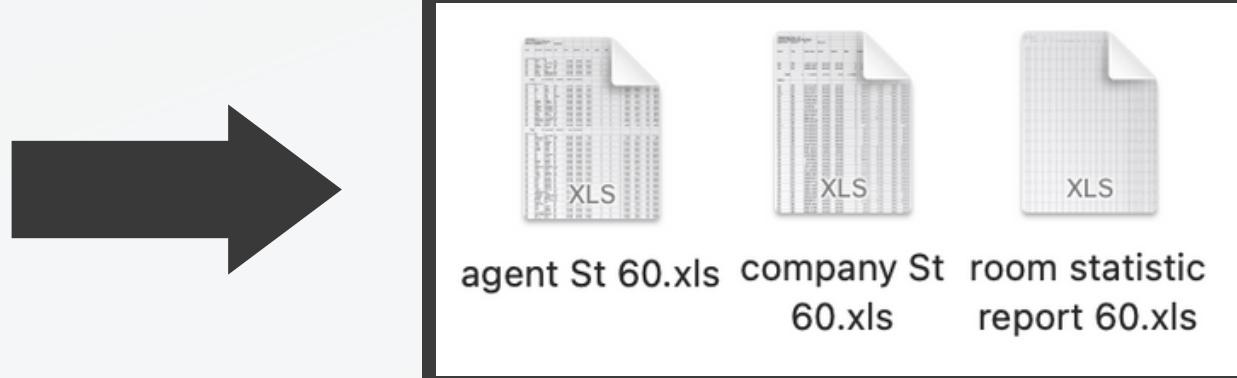
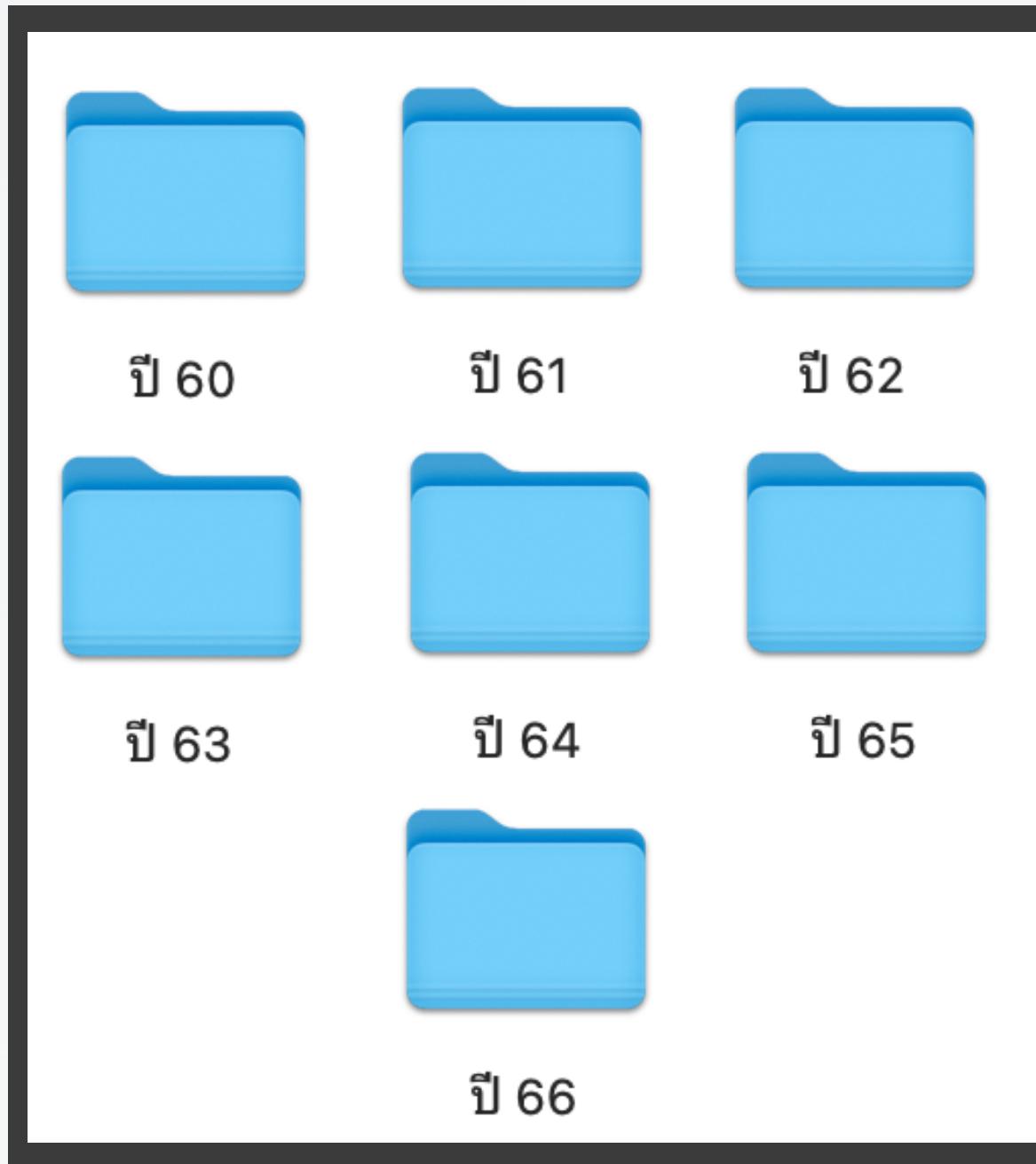
- Able to use the data before Covid-19 pandemic to forecast and calculate the revenue loss in 2020-2021.
- Able to forecast the total revenue of the hotel for the upcoming year in order to visualize the trend to suggest the right business action.
- Able to analyze the result from the model along with from the Exploratory Data Analysis.

02

Data
Collection

RAW DATA (2017-2022)

(EXTRACTED FROM FRONT OFFICE PROGRAM)



3 xls per year
21 files in total

room statistic report.xls

agent St.xls

company St.xls

02

Data
Collection

RAW DATA (2017-2022)

(EXTRACTED FROM FRONT OFFICE PROGRAM)

The screenshot shows a Microsoft Excel spreadsheet titled "room statistic report 60.xls - Compatibility Mode". The spreadsheet contains data from July 30, 2017, to August 21, 2017, for Tubtim Siam Suvarnabhumi Hotel. The data is organized into columns representing various metrics such as Date, Total_Rooms, Guest_Inhouse, Adult, Child, Room_Arrival, Room_Depart, VIP_Guest, Walk_In, CXL, No_Show, Day_Use, Comp., H/U, OO, Occupancy %, Room_Revenue, F&B_Revenue, Others_Revenue, and Total_Revenue. The total revenue for the period is 42,319.16.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1																				
2	Tubtim Siam Suvarnabhumi Hotel																			
3																				
4	ROOM STATISTIC REPORT																			
5	From Date:	30/07/17		To Date:	31/12/17															
6	Date	Total_Rooms	Guest_Inhouse	Adult	Child	Room_Arrival	Room_Depart	VIP_Guest	Walk_In	CXL	No_Show	Day_Use	Comp.	H/U	OO	Occ %	Room_Revenue	F&B_Revenue	Others_Revenue	Total_Revenue
7	30/07/17	37	37	69	0	37	0	0	0	0	0	0	0	0	0	100.00	33,559.16	8,760.00	0.00	42,319.16
11	31/07/17	37	35	69	1	19	21	0	0	0	0	0	0	0	0	94.59	22,410.00	5,580.00	0.00	27,990.00
13	01/08/17	47	32	63	2	19	22	1	1	0	0	0	0	0	0	68.09	16,080.00	5,700.00	0.00	21,780.00
14	02/08/17	47	34	67	0	17	15	1	1	0	0	0	0	0	0	72.34	18,351.25	6,030.00	0.00	24,381.25
15	03/08/17	47	38	76	0	24	20	1	0	0	0	1	0	1	0	80.85	18,730.00	6,660.00	0.00	25,390.00
16	04/08/17	47	11	20	1	4	31	0	1	0	0	0	0	0	0	23.40	6,535.31	1,845.00	0.00	8,380.31
17	05/08/17	47	25	47	2	23	9	0	3	0	0	0	0	0	0	53.19	18,740.18	2,880.00	0.00	21,620.18
18	06/08/17	47	38	75	0	21	8	0	1	0	0	0	0	0	0	80.85	23,700.00	4,020.00	40.00	27,760.00
19	07/08/17	47	23	47	0	2	17	0	1	0	0	0	0	0	0	48.94	13,150.00	4,230.00	0.00	17,380.00
20	08/08/17	47	27	54	0	25	21	0	0	0	0	3	0	0	0	57.45	24,382.60	4,860.00	600.00	29,842.60
21	09/08/17	47	28	55	2	27	26	0	2	0	1	0	0	0	0	59.57	18,420.41	4,950.00	0.00	23,370.41
22	10/08/17	47	39	78	2	36	25	0	2	0	0	0	0	0	0	82.98	26,094.30	7,110.00	-500.00	32,704.30
23	11/08/17	47	17	34	2	6	28	0	2	0	0	0	0	0	0	36.17	10,527.00	2,970.00	0.00	13,497.00
24	12/08/17	47	29	68	0	29	17	0	0	0	0	0	0	0	0	61.70	20,913.79	5,280.00	0.00	26,193.79
25	13/08/17	47	29	58	0	14	14	0	1	0	0	0	0	0	0	61.70	21,439.42	5,040.00	0.00	26,479.42
26	14/08/17	47	38	73	0	24	15	1	1	0	0	0	0	0	0	80.85	33,056.50	0.00	0.00	33,056.50
27	15/08/17	47	19	37	0	3	22	1	0	0	0	0	0	0	0	40.43	16,927.58	3,530.00	0.00	20,457.58
28	16/08/17	47	34	68	0	19	4	0	0	0	3	0	0	0	0	72.34	26,150.00	5,760.00	0.00	31,910.00
29	17/08/17	47	20	38	2	5	19	2	0	0	0	0	0	0	0	42.55	14,276.52	2,700.00	0.00	16,976.52
30	18/08/17	47	2	4	0	1	19	0	0	0	0	0	0	0	0	4.26	2,200.00	420.00	0.00	2,620.00
31	19/08/17	47	3	6	0	2	1	0	0	0	0	0	0	0	0	6.38	1,916.52	360.00	0.00	2,276.52
32	20/08/17	47	1	2	0	1	3	0	0	0	0	0	0	0	0	2.13	1,110.00	540.00	0.00	1,650.00
33	21/08/17	47	6	12	0	6	1	0	0	0	0	0	0	0	0	12.77	7,201.00	900.00	0.00	8,101.00

room statistic report.xls

The daily record of total revenue that gained from each year

02

Data
Collection

RAW DATA (2017-2022)

(EXTRACTED FROM FRONT OFFICE PROGRAM)

The screenshot shows a Microsoft Excel spreadsheet titled "agent St 60.xls - Compatibility Mode". The spreadsheet contains data from 2017 to 2022, extracted from a front office program. The data is organized into several sections:

- Header:** Titled "Tubtim Siam Page -1 of 1" and "Agent Statistic Report By Detail". It includes fields for "Report from" (30/07/17) and "To" (31/12/17).
- Room Types:** A section listing room types and their descriptions.
- AGODA COMPANY PTE.LTD.:** A large section containing 18 rows of guest information. Each row includes columns for Room, Last name, First name, Title, Arrival, Departure, Rate, Nights, Adult, Child, Room Rev., F&B Rev., Other Rev., and Total Rev. Row 18 shows a total for AGODA.
- AT NETWORK CO.,LTD.:** A section containing 42 rows of guest information, similar to the AGODA section but with different guest names and room details.
- Total Rows:** The spreadsheet has 44 rows in total, with the last row being a summary.

agent St.xls

The daily record of total
revenue separated by each
agent in each year

02Data
Collection

RAW DATA (2017-2022)

(EXTRACTED FROM FRONT OFFICE PROGRAM)

	A	B	C	D	E	F	G	H	I	J
1	A	B	C	D	E	F	G	H	I	J
2	Tubtim Siam Su	Page -1 of 1								
3	Company Statistic Report By Detail									
4	Report from	30/07/17	To	30/12/17						
5										
6										
7	Room	Type	Guest name	Arrival	Departl	Night	Room Revenue	F&B Revenue	Others Revenue	Revenue Total
8										
9										
10										
11	622	TBR	กัลป์ นิ่ม อรุณรัตน์	18/11/17	19/11/17	1	610.00	180.00	0.00	790.00
12	404	DLX	กัลป์ อรุณรัตน์	06/12/17	09/12/17	2	3,310.00	270.00	0.00	3,580.00
13	403	DLX	กัลป์ อรุณรัตน์	06/12/17	09/12/17	2	2,400.00	180.00	0.00	2,580.00
14										
15	Total	5	6,320.00	630.00	0.00	6,950.00				
16										
17	Call-in									
18										
19	622	TBR	Chen Jing Zhang Z.	07/10/17	14/10/17	6	3,000.00	0.00	0.00	3,000.00
20	505	DLX	Chen pei-Tsen Ms.	08/12/17	09/12/17	1	1,200.00	90.00	0.00	1,290.00
21	506	DLX	Chen Yi-Chen Ms.	08/12/17	09/12/17	1	1,200.00	90.00	0.00	1,290.00
22	Z013	ZZZ	China Group (MA	07/10/17	23/10/17	0	38,000.00	0.00	400.00	38,400.00
23	621	TBR	Deng Hao Li Bingr.	07/10/17	14/10/17	6	3,000.00	0.00	0.00	3,000.00
24	632	TBR	He Xikai Yang Lin	07/10/17	14/10/17	6	3,000.00	0.00	0.00	3,000.00
25	401	DLX	Hisashi Hoshide M	28/12/17	29/12/17	1	1,245.00	45.00	0.00	1,290.00
26	625	TBR	Li Songai Sun Limi	07/10/17	14/10/17	6	3,000.00	0.00	0.00	3,000.00
27	503	DLX	Ling Wei Ms.	28/12/17	31/12/17	3	3,735.00	135.00	0.00	3,870.00

company St.xls

The daily record of total revenue separated by each company in each year. (Noted that check-in channels such as walk-in, LineOfficial is also counted as one of the company in this file)

03

Data
Preprocessing

CLEAN DATA

Implement a function that **iterate** through every files in the specific folder and **concat** all of them as 1 dataframe



Iterate through the rows to find the specific row that needed to be used as the real column name



Implement a function that get the dataframe of the specific agent/company that we want from the concatenated dataframe from the previous step by **iterating** over the start index matching with the end index of the word "Total". Then **concat** all of them in 1 separate dataframe based on each specified agent/company name.



- **Drop** the columns that has all NaN values
- **Impute** some values in specific column
- **Create** new calculated /categories column for doing an EDA

04

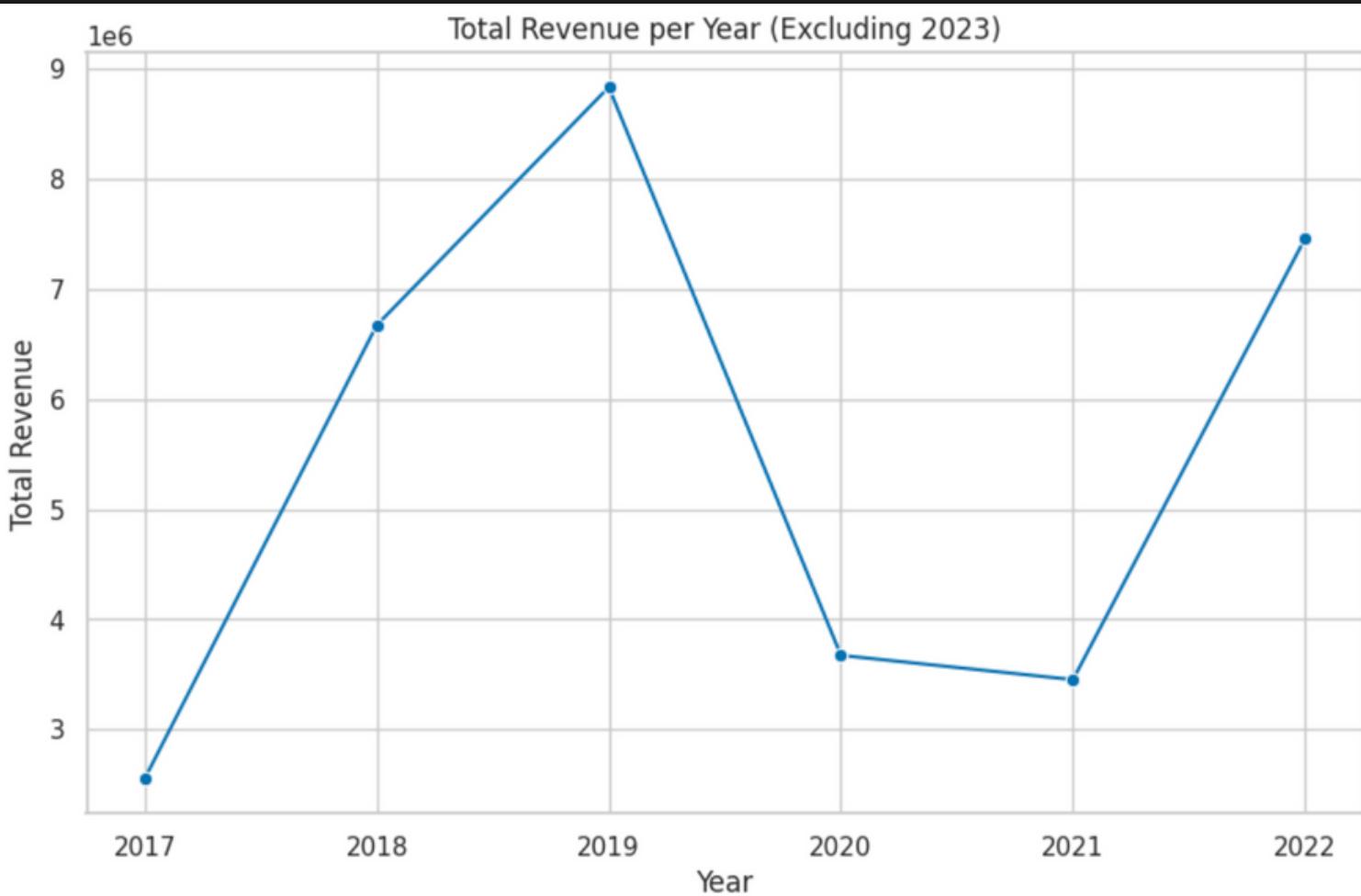
Exploratory
Data
Analysis

TOTAL REVENUE

Total Revenue per Day

Statistic	Value
count	2090
mean	16767.4
std	9729.86
min	0
25%	9470.02
50%	16264.2
75%	22817.2
max	65805.6

Total Revenue per Year (2017-2022)



- The average total revenue per day of this hotel is **\$16,767**. Moreover, this hotel is able to reach the maximum of **\$65,805** in one day.
- The total revenue is on an **increasing trend** during 2017-2019. However, Covid-19 pandemic caused a **huge drop** in total revenue.
- Currently, although Covid situation has been better, the total revenue in **2022** still has **NOT** been recovered.

04

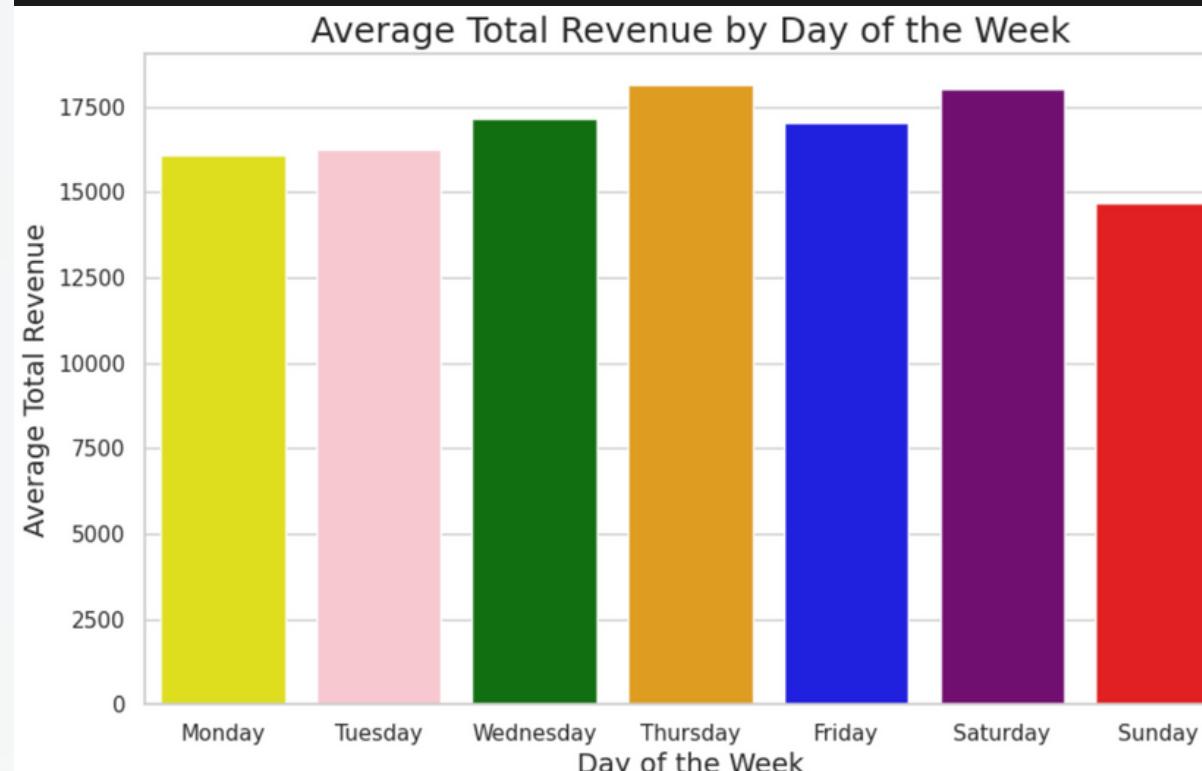
Exploratory
Data
Analysis

WEEKLY SALES

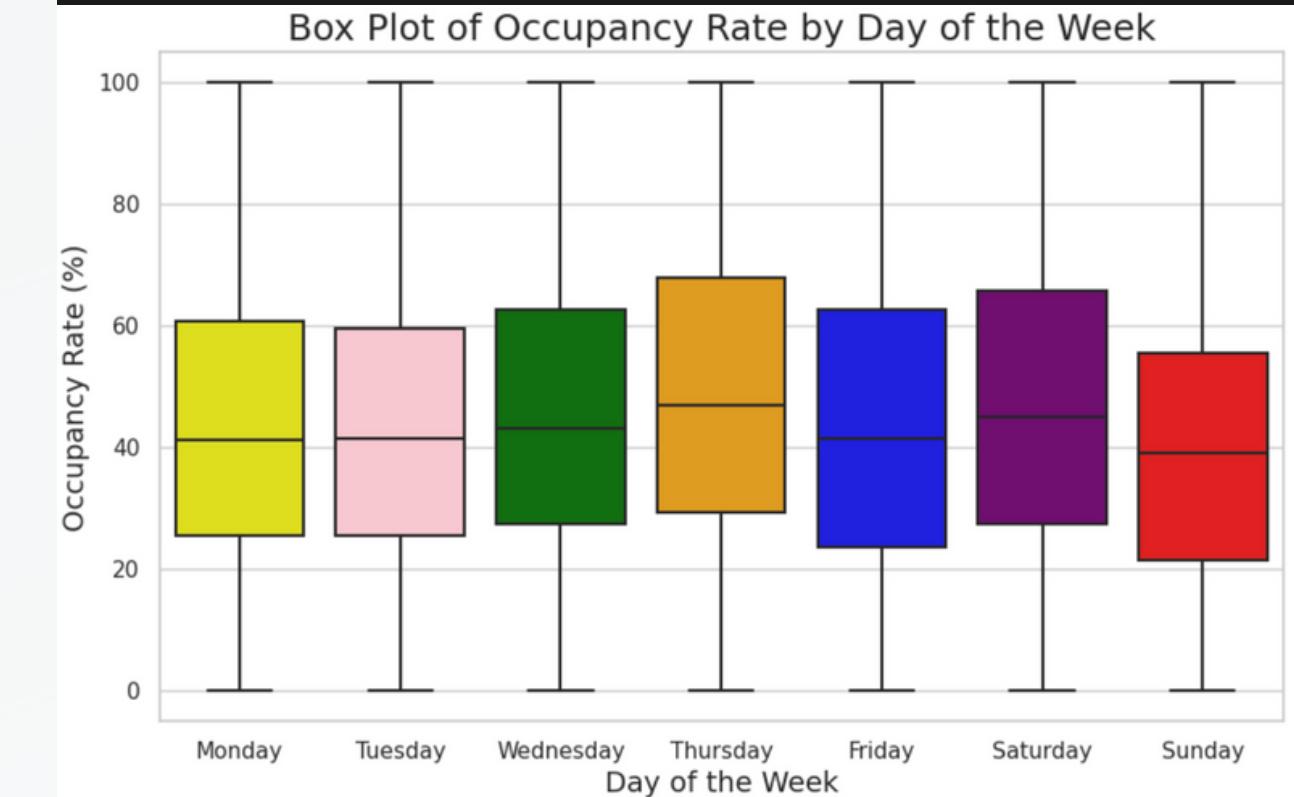
Average Total Revenue in Each Day of the Week

Day of Week	Average Total Revenue
Monday	16098.6
Tuesday	16242.6
Wednesday	17149
Thursday	18130.4
Friday	17041.3
Saturday	18022.9
Sunday	14697

Average Total Revenue by Day of the Week



Occupancy Rate by Day of the Week



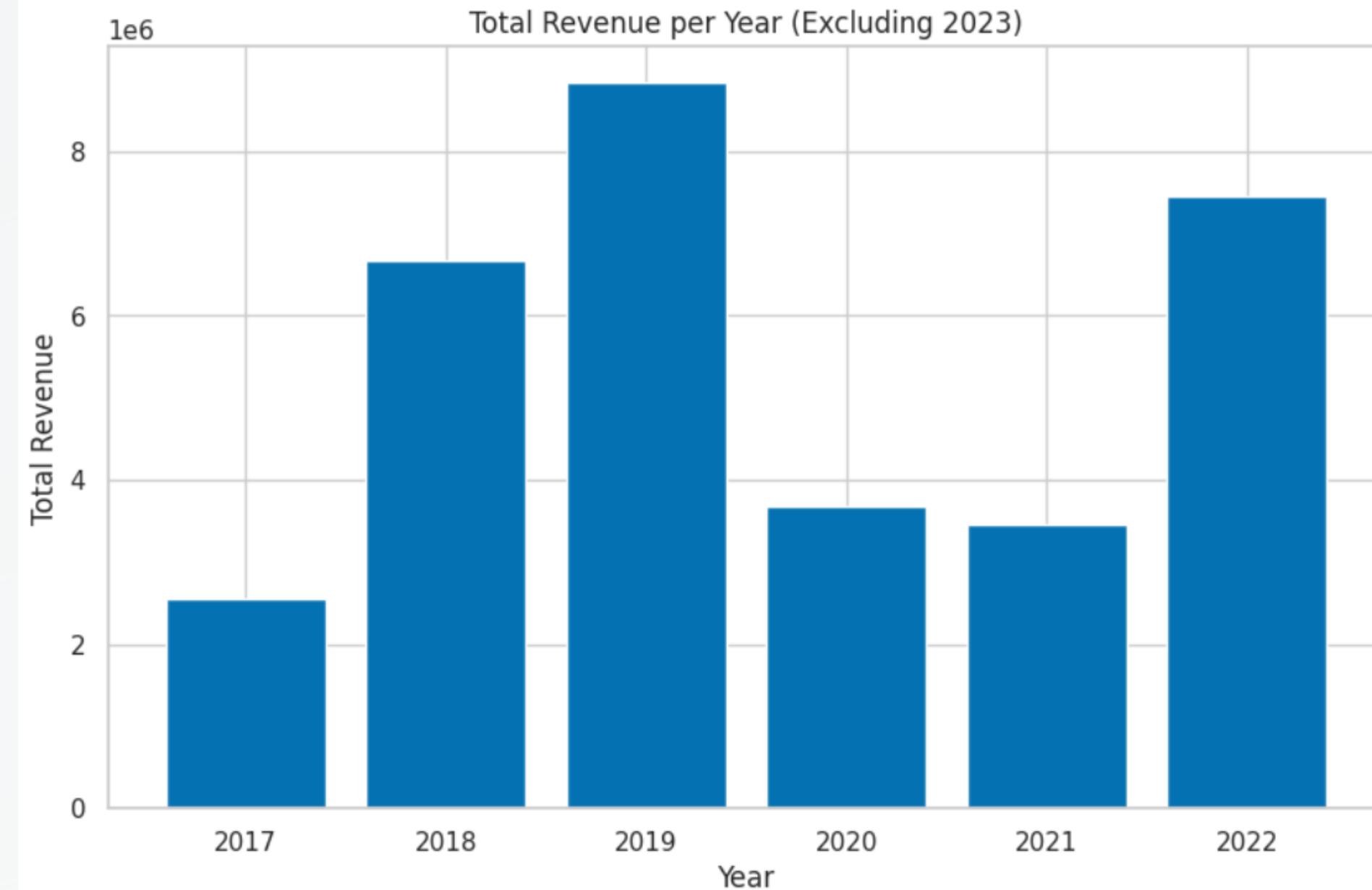
- It can be observed that **Sunday** earned the **lowest** average **total revenue** and **occupancy rate** among all other days. This is because this hotel is a business hotel, so customers tend to book more for the rooms during the weekdays for meetings & working rather than on the holidays.
- In order to boost the revenue on Sunday, the hotel should target other group of customer rather than business person, such as travelers

04

Exploratory
Data
Analysis

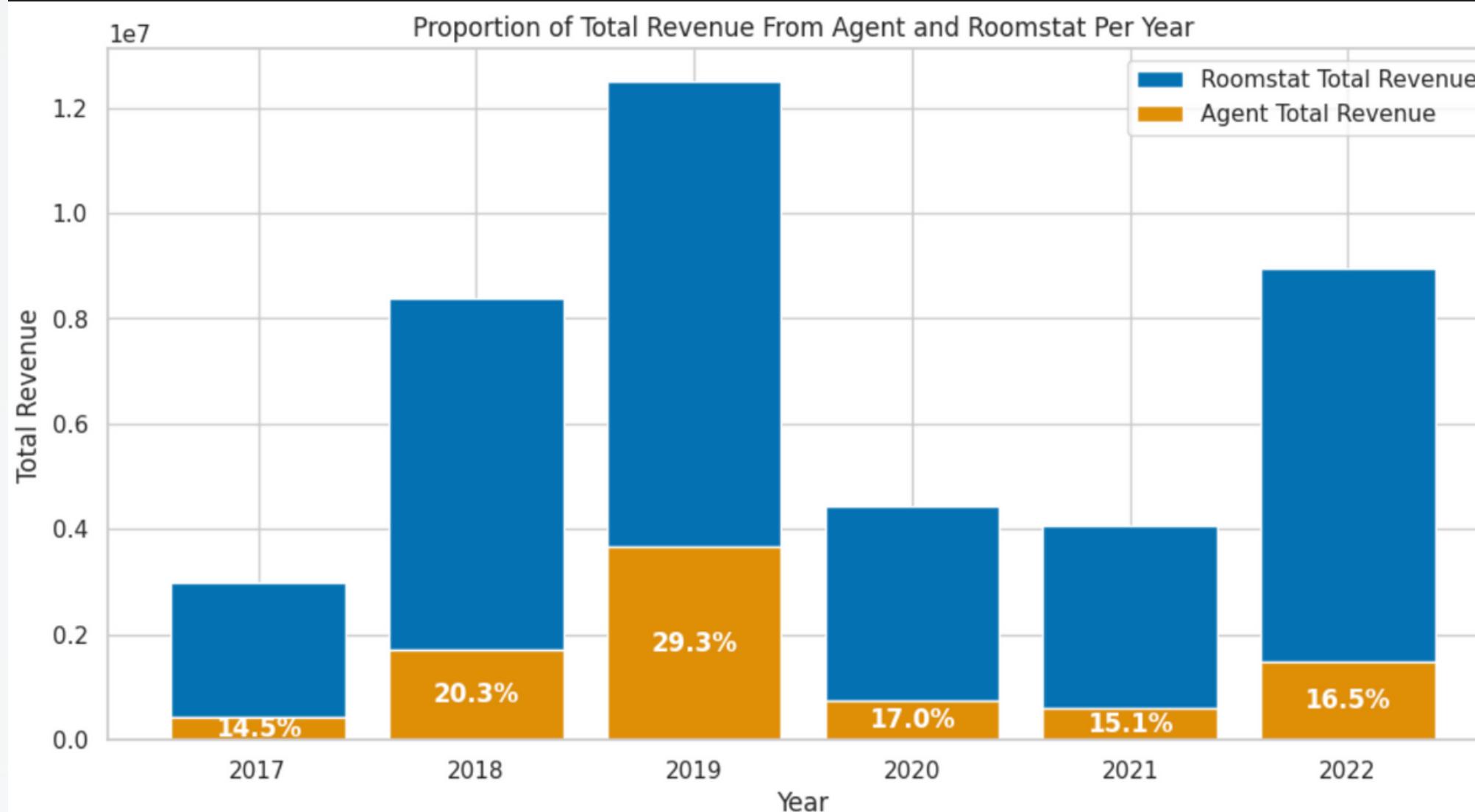
TOTAL REVENUE

Total Revenue per Year



REVENUE STRUCTURE

Proportion of Revenue from Agent in Total Revenue



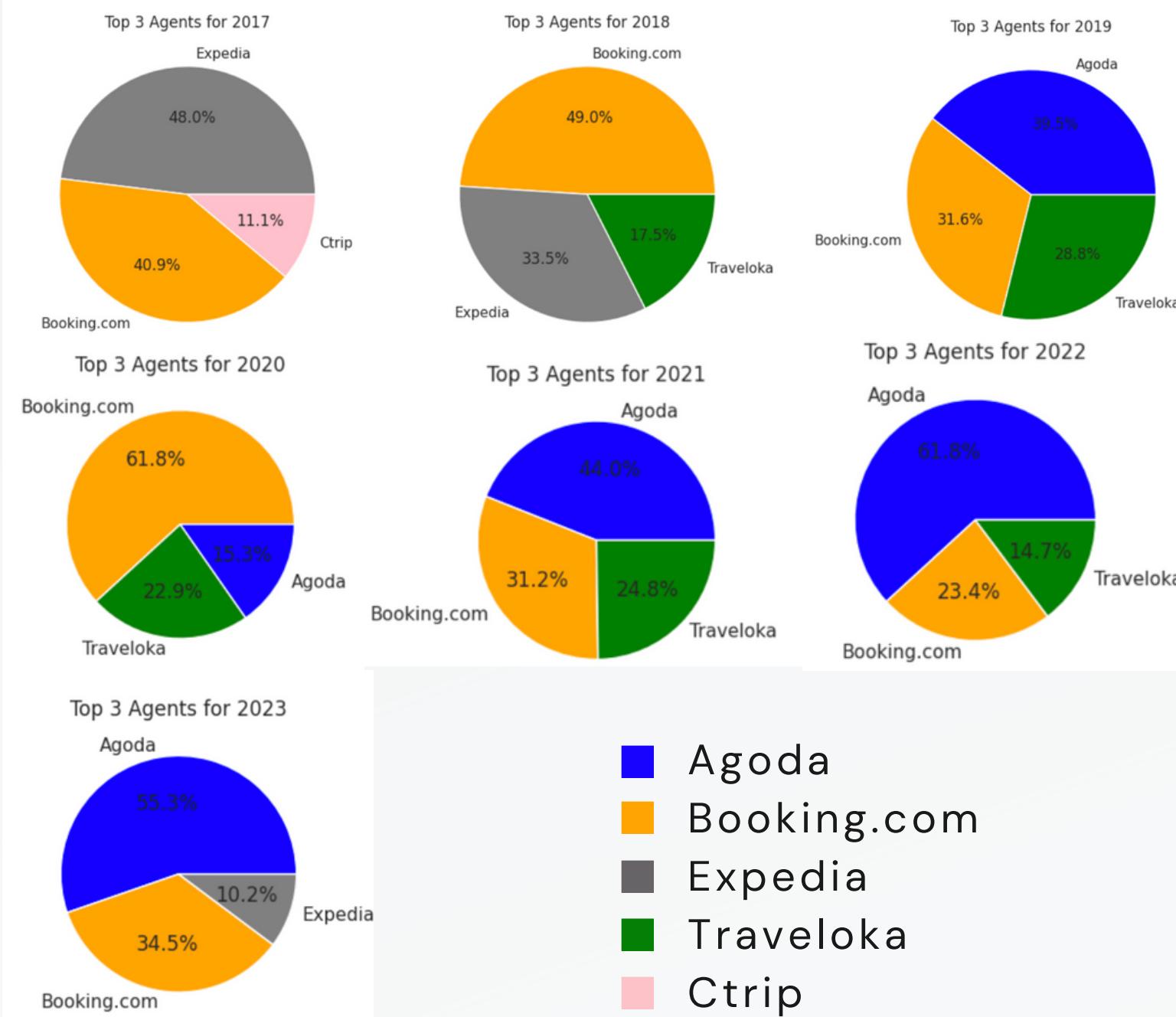
- It can be observed that Online Travel Agency (OTA) is one of the **important factors** to boost the yearly total revenue.
- Based on the graph, the percentage of revenue from OTA is **positively correlated** with total revenue, which infer that it is **worth investing in OTA** in order to boost the total revenue.

04

Exploratory
Data
Analysis

ONLINE TRAVEL AGENCY (OTA) SALES

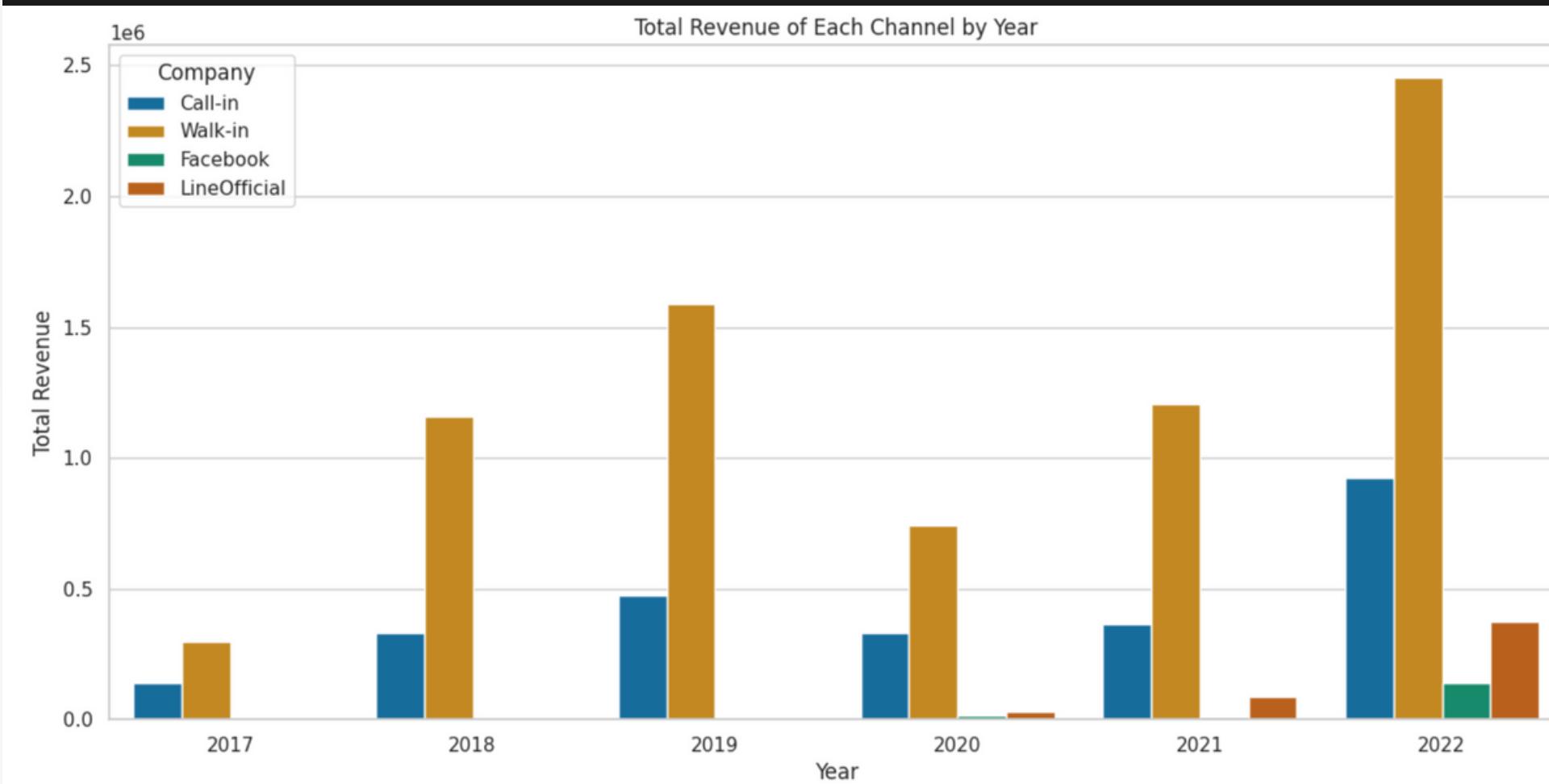
Top 3 OTAs in Each Year



- Expedia was once an agency that created the highest total revenue for the hotel during the **first two years** since the hotel opened.
- Agoda has become **#1 agency** that created the highest total revenue for several years except in 2020 that has Covid-19 pandemic, the revenue that hotel gained from Agoda has **significantly dropped**. This is because most of the customers that came from Agoda are **foreigners**.
- Booking.com has always on **top 2 agencies** that create the highest total revenue to the hotel. During Covid-19 pandemic, the revenue from booking.com was **still high**. This is because most of the customers that came from Booking.com are **local people**.
- So, it is **worth** investing on Agoda and Booking.com

NON-OTA SALES

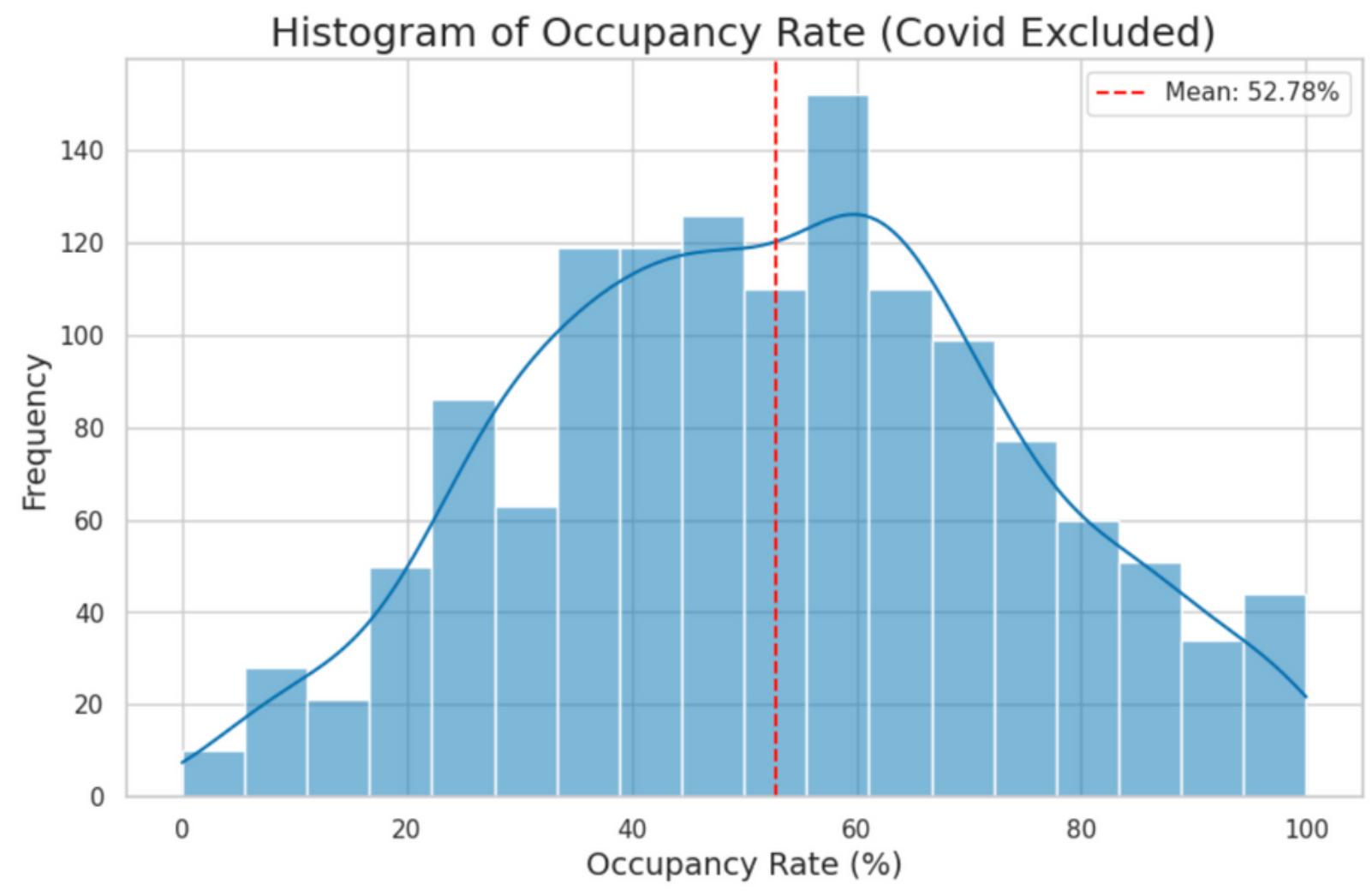
Total Revenue From Each Check-in Channel by Year



- Among all these 4 types of check-in channels, the highest total revenue is gained from the customer that **walk-in** to the hotel.
- It can be presumed that the hotel is located on the area that people can find the location of the hotel, which is near the airport, via **GPS application**, ex: Google Map, and can easily walk in. So, it should be **worth trying to invest** on those platforms.
- It can be observed that after Covid-19 pandemic, Line and Facebook Official are used to **boost** the walk-in and call-in rate.

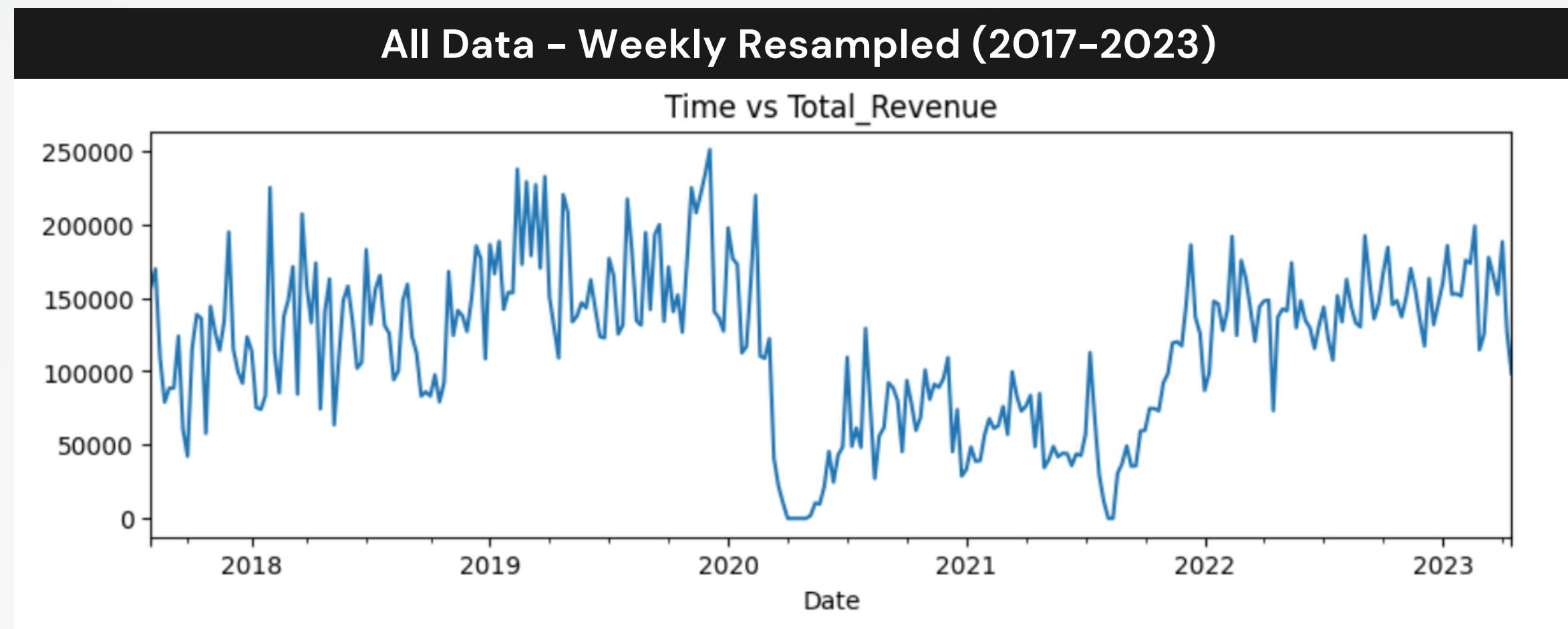
OCCUPANCY RATE

Occupancy Rate (Covid Excluded)



- It can be observed that the average percentage of occupancy rate (Occ%) is **52.78%** during 2017–2022, covid excluded,

TIME SERIES

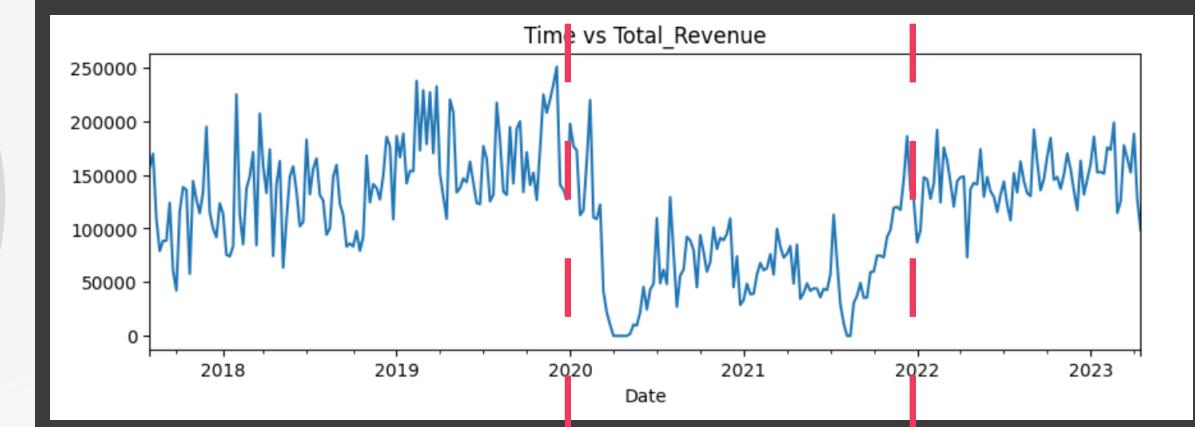


- It can be observed that that COVID-19 pandemic caused a huge impact on hotel's total revenue during **2020** to **late 2021**
- So, The modeling will be done by 3 range of date

05

Modeling

TIME SERIES



BEFORE COVID
(2017-2019)

AFTER COVID
(2022-2023)

ALL DATA
(2017-2023)

Daily Data

- Preprocessing
- Stationary test
- AR Model
- ARMA Model

Weekly Data

- Preprocessing
- Stationary test
- SARIMA Model
- Future Forecast
- Loss Comparison

Daily Data

- Preprocessing
- Stationary test
- ARMA Model

Weekly Data

NOT ENOUGH DATA (around 65)
TO TRAIN

Daily Data

- Preprocessing
- Stationary test
- ARMA Model

Weekly Data

- Preprocessing
- Stationary test
- SARIMA Model
- Future Forecast

Noted that weekly data is the data that is
result_df.resample('W', on='Date').sum()

05

Modeling

TIME SERIES

BEFORE COVID
(2017-2019)

AFTER COVID
(2022-2023)

Weekly Data

- Less Noise
- Less MAPE Error

ALL DATA
(2017-2023)

✗ Daily Data ✓ Weekly Data

- Preprocessing
- Stationary test
- AR Model
- ARMA Model

- Preprocessing
- SARIMA Model
- Future Forecast
- Loss Comparation

✗ Daily Data

- Preprocessing
- Stationary test
- ARMA Model

✗ Weekly Data

NOT ENOUGH DATA (around 65)
TO TRAIN

✗ Daily Data

- Preprocessing
- Stationary test
- ARMA Model

✓ Weekly Data

- Preprocessing
- Stationary test
- SARIMA Model
- Future Forecast

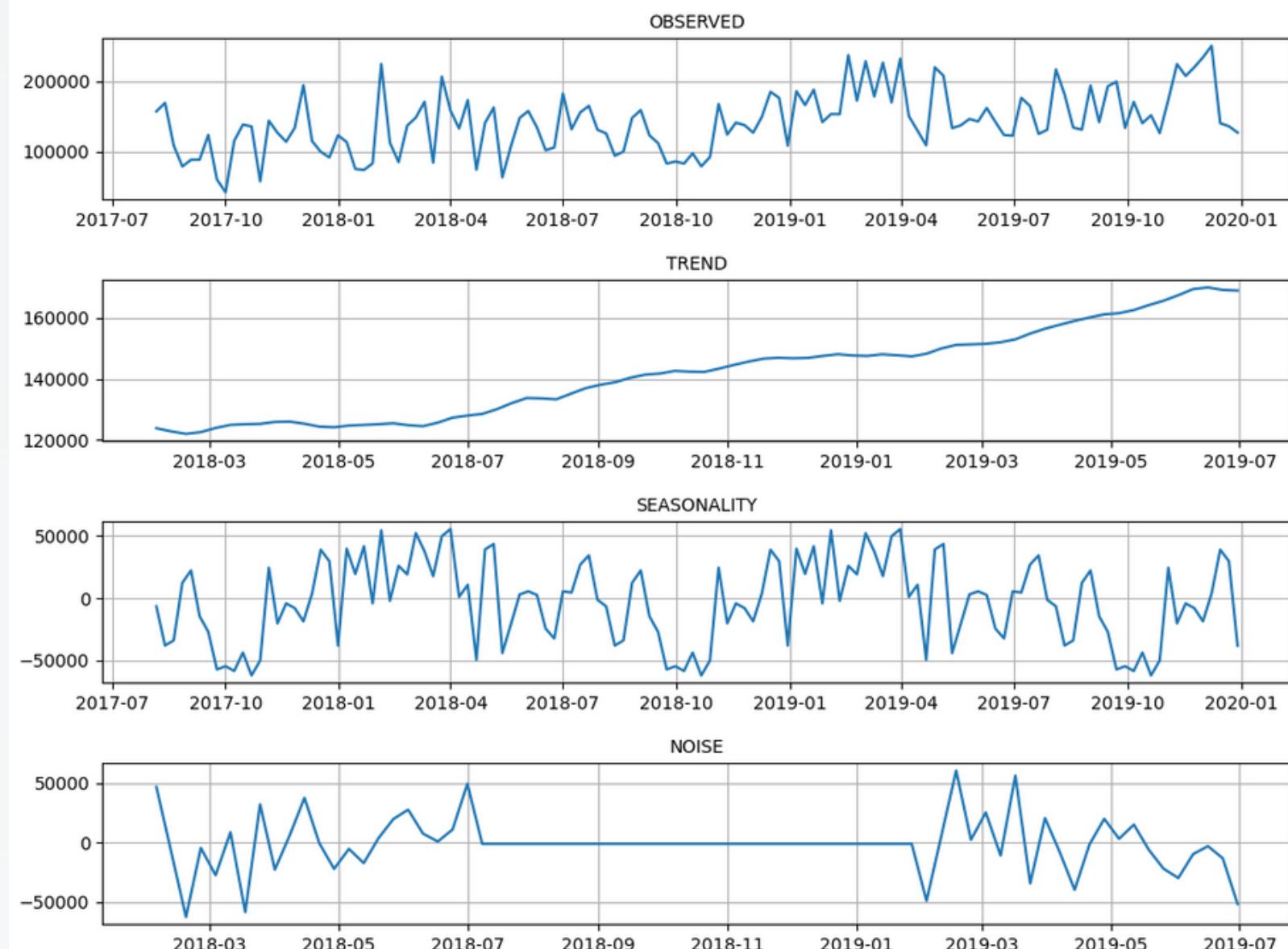
05

Modeling

BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
(2017-08-06 TO 2019-12-29)

`seasonal_decomposed(period = 52)`



Trend

The linear trend **can be observed**

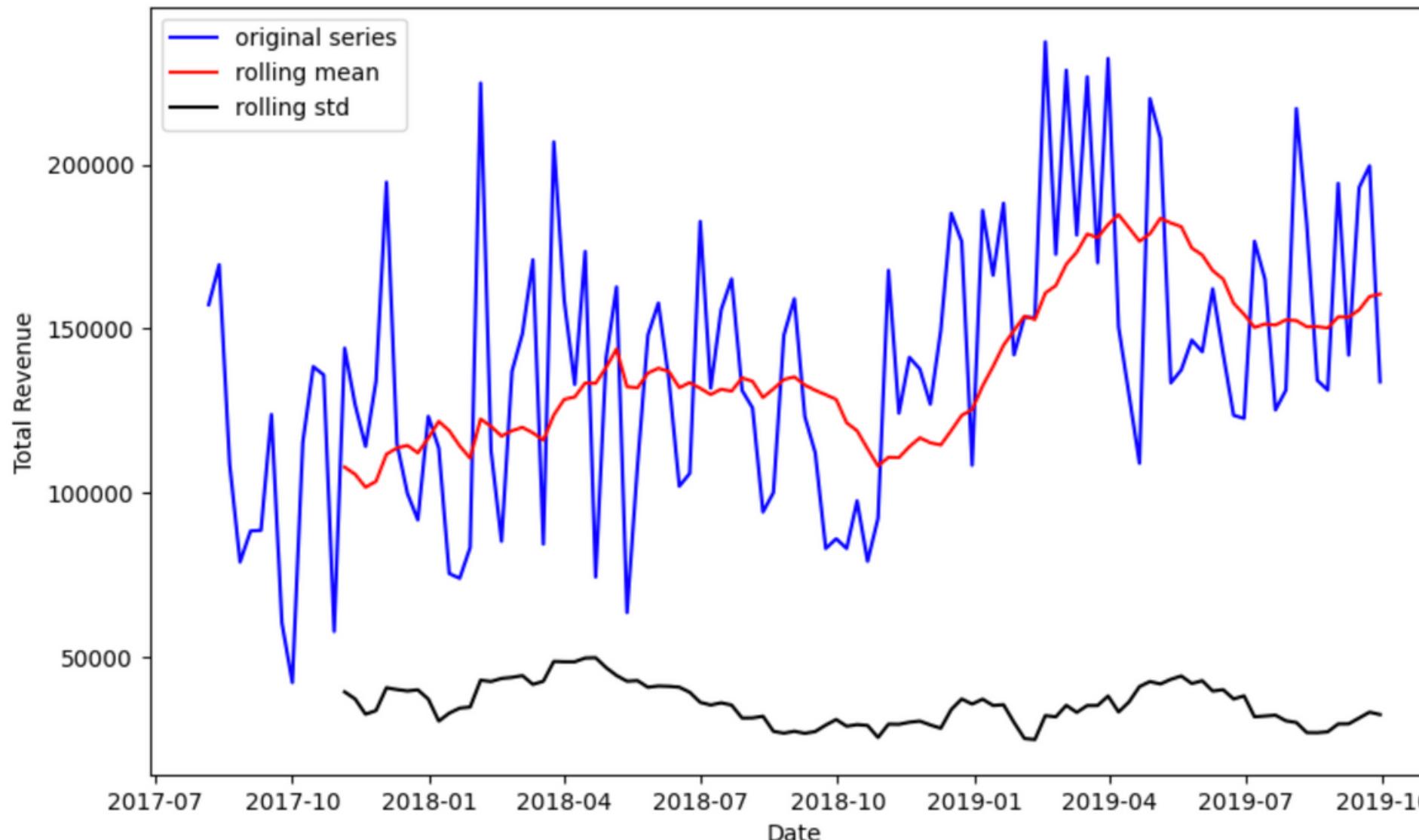
Seasonality

Seasonality can be **barely observed** but
not too obvious

BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
(2017-08-06 TO 2019-12-29)

Rolling mean & std



- It can be observed that the rolling mean is **not stationary**
- From the ADF test, p-value of the data is **0.091**, which is > 0.05 that implies the **non-stationary**. So, I decide to use **SARIMA model**

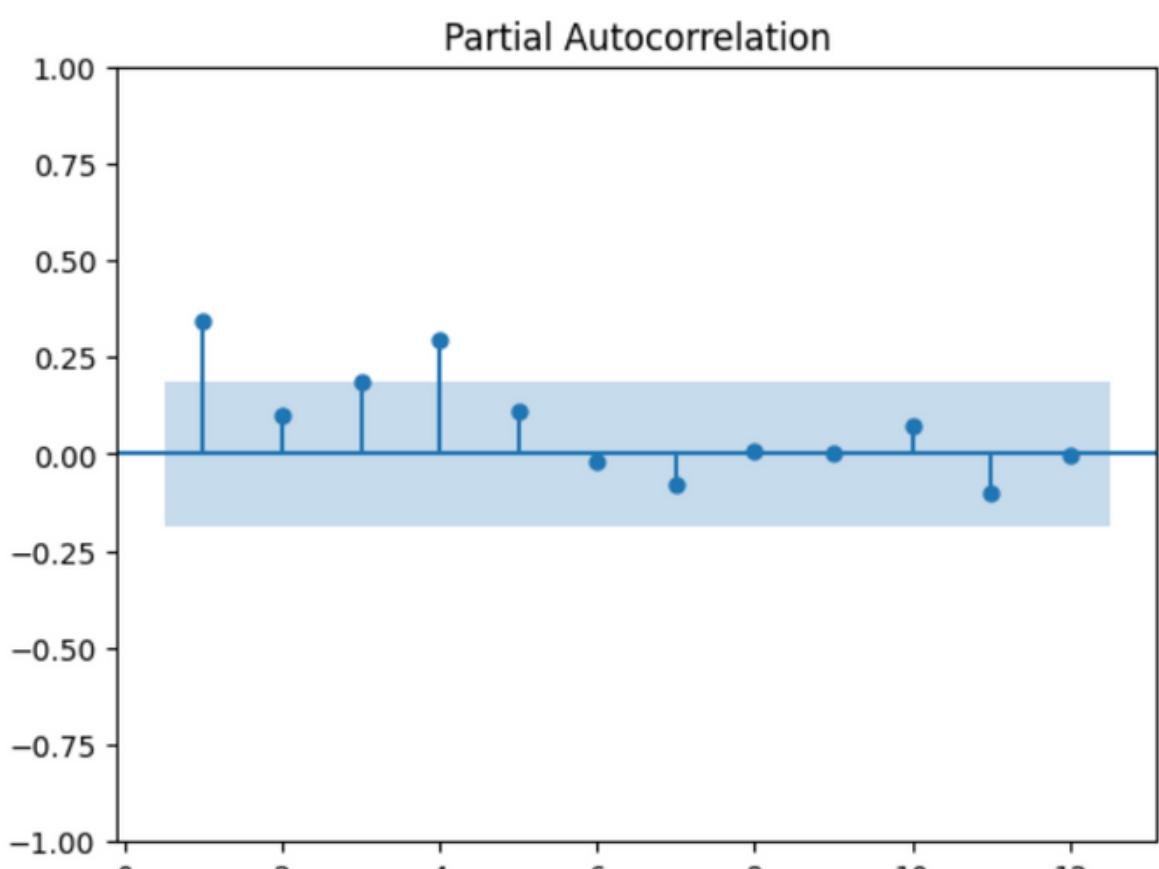
05

Modeling

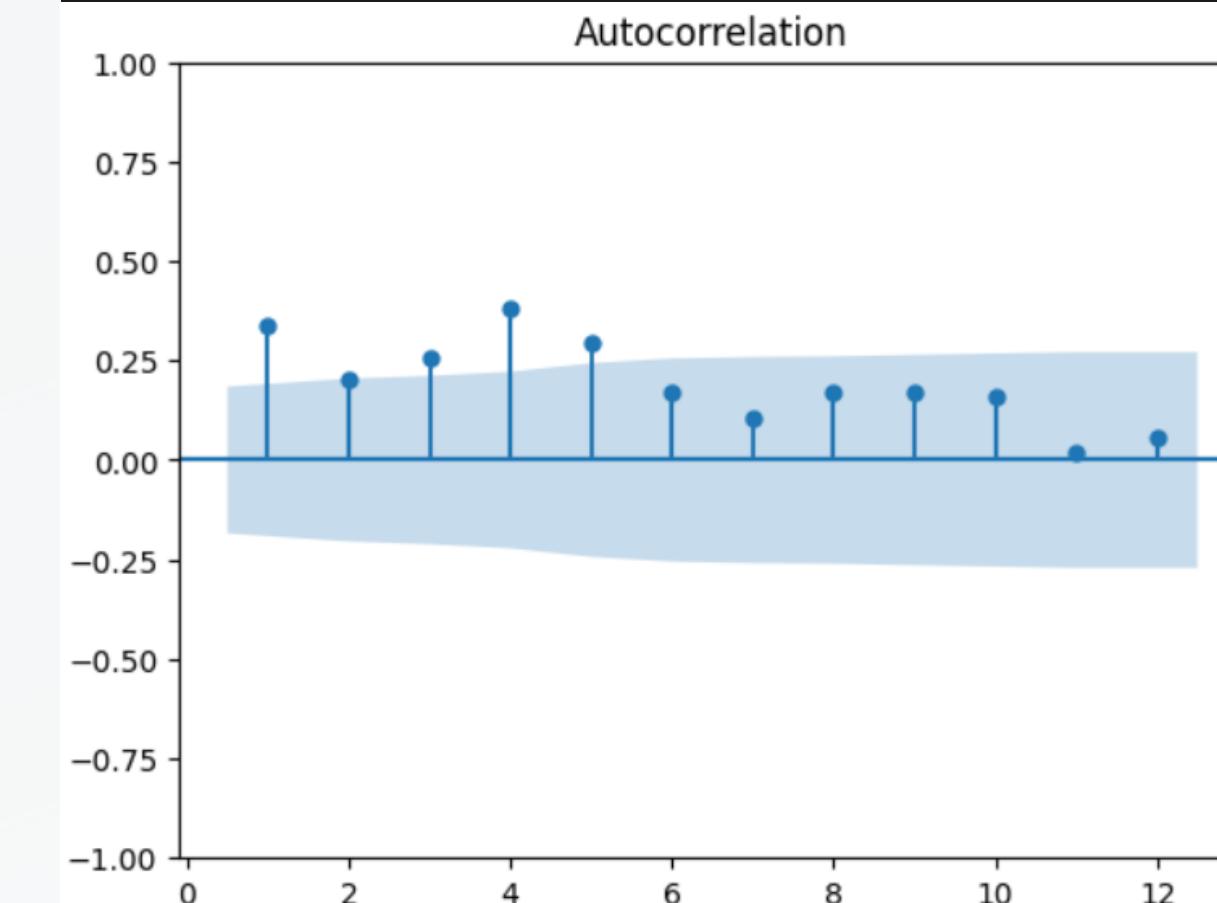
BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
(2017-08-06 TO 2019-12-29)

PACF



ACF



$p = 4$

$q = 10$

05

Modeling

BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
 (2017-08-06 TO 2019-12-29)

order = (4,1,10), seasonal_order =(1,1,0,52)

SARIMAX Results

Dep. Variable:	Total_Revenue	No. Observations:	113		
Model:	SARIMAX(4, 1, 10)x(1, 1, [], 52)	Log Likelihood	-709.165		
Date:	Thu, 27 Apr 2023	AIC	1450.330		
Time:	11:04:44	BIC	1483.840		
Sample:	08-06-2017 - 09-29-2019	HQIC	1463.438		
Covariance Type:	opg				
coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.4825	0.190	-2.538	0.011 -0.855	-0.110
ar.L2	-0.3184	0.163	-1.951	0.051 -0.638	0.001
ar.L3	-0.5672	0.192	-2.959	0.003 -0.943	-0.192
ar.L4	-0.3287	0.189	-1.738	0.082 -0.699	0.042
ma.L1	0.0926	0.290	0.319	0.750 -0.477	0.662
ma.L2	-0.0910	0.207	-0.440	0.660 -0.496	0.314
ma.L3	0.2112	0.210	1.007	0.314 -0.200	0.622
ma.L4	0.2702	0.215	1.257	0.209 -0.151	0.692
ma.L5	-0.4663	0.126	-3.711	0.000 -0.713	-0.220
ma.L6	-0.1062	0.154	-0.689	0.491 -0.408	0.196
ma.L7	0.1111	0.192	0.579	0.562 -0.265	0.487
ma.L8	0.0633	0.195	0.325	0.745 -0.318	0.445
ma.L9	-0.1450	0.206	-0.705	0.481 -0.548	0.258
ma.L10	0.6491	0.210	3.088	0.002 0.237	1.061
ar.S.L52	-0.8852	0.054	-16.352	0.000 -0.991	-0.779
sigma2	7.824e+08	1.05e-09	7.43e+17	0.000 7.82e+08	7.82e+08
Ljung-Box (L1) (Q):	0.04	Jarque-Bera (JB):	1.14		
Prob(Q):	0.85	Prob(JB):	0.57		
Heteroskedasticity (H):	1.09	Skew:	-0.01		
Prob(H) (two-sided):	0.84	Kurtosis:	2.33		

order = (3,1,10), seasonal_order =(1,1,0,52)

SARIMAX Results

Dep. Variable:	Total_Revenue	No. Observations:	113		
Model:	SARIMAX(3, 1, 10)x(1, 1, [], 52)	Log Likelihood	-711.345		
Date:	Thu, 27 Apr 2023	AIC	1452.690		
Time:	11:02:05	BIC	1484.105		
Sample:	08-06-2017 - 09-29-2019	HQIC	1464.978		
Covariance Type:	opg				
coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.3100	0.148	-2.096	0.036 -0.600	-0.020
ar.L2	-0.4131	0.104	-3.953	0.000 -0.618	-0.208
ar.L3	-0.9218	0.144	-6.402	0.000 -1.204	-0.640
ma.L1	-0.2210	0.169	-1.307	0.191 -0.552	0.110
ma.L2	0.1615	0.162	0.994	0.320 -0.157	0.480
ma.L3	0.4927	0.193	2.552	0.011 0.114	0.871
ma.L4	-0.2719	0.185	-1.472	0.141 -0.634	0.090
ma.L5	-0.3224	0.157	-2.053	0.040 -0.630	-0.015
ma.L6	-0.0346	0.144	-0.241	0.810 -0.316	0.247
ma.L7	0.2675	0.167	1.605	0.108 -0.059	0.594
ma.L8	-0.0482	0.149	-0.324	0.746 -0.339	0.243
ma.L9	0.0031	0.144	0.022	0.983 -0.278	0.284
ma.L10	0.2842	0.119	2.382	0.017 0.050	0.518
ar.S.L52	-0.7910	0.060	-13.109	0.000 -0.909	-0.673
sigma2	7.916e+08	8.68e-11	9.12e+18	0.000 7.92e+08	7.92e+08
Ljung-Box (L1) (Q):	0.06	Jarque-Bera (JB):	0.90		
Prob(Q):	0.81	Prob(JB):	0.64		
Heteroskedasticity (H):	1.16	Skew:	0.16		
Prob(H) (two-sided):	0.75	Kurtosis:	2.49		

In can be observed that the order of (4,1,10) from the previous slide results in getting the p-value of ar.L4 = 0.082, which is > 0.05. So, the solution is to stepping the order down to (3,1,0) and check the p-value again, which ar.L3 and ma.L10 is now both < 0.05, which is now **significant**



05

Modeling

BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
 (2017-08-06 TO 2019-12-29)

order = (3,1,10), seasonal_order = (1,1,0,52) , trend = 't'						
SARIMAX Results						
Dep. Variable:		Total_Revenue				
Model:		SARIMAX(3, 1, 10)x(1, 1, 0, 52)				
Date:		Thu, 27 Apr 2023				
Time:		11:07:07				
Sample:		08-06-2017 - 09-29-2019				
Covariance Type: opg						
coef	std err	z	P> z	[0.025	0.975]	
drift	107.8840	80.100	1.347	0.178	-49.108	264.876
ar.L1	-0.3151	0.128	-2.454	0.014	-0.567	-0.063
ar.L2	-0.4161	0.093	-4.483	0.000	-0.598	-0.234
ar.L3	-0.9274	0.131	-7.104	0.000	-1.183	-0.671
ma.L1	-0.1738	0.157	-1.104	0.270	-0.482	0.135
ma.L2	0.1831	0.141	1.294	0.195	-0.094	0.460
ma.L3	0.5401	0.171	3.152	0.002	0.204	0.876
ma.L4	-0.1980	0.163	-1.212	0.225	-0.518	0.122
ma.L5	-0.2780	0.148	-1.884	0.060	-0.567	0.011
ma.L6	-0.0090	0.131	-0.068	0.945	-0.266	0.248
ma.L7	0.3603	0.161	2.244	0.025	0.046	0.675
ma.L8	-0.0264	0.145	-0.182	0.856	-0.311	0.258
ma.L9	0.0186	0.130	0.143	0.886	-0.236	0.273
ma.L10	0.3281	0.129	2.546	0.011	0.075	0.581
ar.S.L52	-0.7567	0.057	-13.302	0.000	-0.868	-0.645
sigma2	8.149e+08	1.21e-05	6.74e+13	0.000	8.15e+08	8.15e+08
Ljung-Box (L1) (Q):	0.34	Jarque-Bera (JB):	0.91			
Prob(Q):	0.56	Prob(JB):	0.64			
Heteroskedasticity (H):	1.34	Skew:	0.16			
Prob(H) (two-sided):	0.52	Kurtosis:	2.49			



order = (3,1,10), seasonal_order = (1,1,0,52) , trend = 'n'						
SARIMAX Results						
Dep. Variable:		Total_Revenue				
Model:		SARIMAX(3, 1, 10)x(1, 1, 0, 52)				
Date:		Thu, 27 Apr 2023				
Time:		11:02:05				
Sample:		08-06-2017 - 09-29-2019				
Covariance Type: opg						
coef	std err	z	P> z	[0.025	0.975]	
ar.L1	-0.3100	0.148	-2.096	0.036	-0.600	-0.020
ar.L2	-0.4131	0.104	-3.953	0.000	-0.618	-0.208
ar.L3	-0.9218	0.144	-6.402	0.000	-1.204	-0.640
ma.L1	-0.2210	0.169	-1.307	0.191	-0.552	0.110
ma.L2	0.1615	0.162	0.994	0.320	-0.157	0.480
ma.L3	0.4927	0.193	2.552	0.011	0.114	0.871
ma.L4	-0.2719	0.185	-1.472	0.141	-0.634	0.090
ma.L5	-0.3224	0.157	-2.053	0.040	-0.630	-0.015
ma.L6	-0.0346	0.144	-0.241	0.810	-0.316	0.247
ma.L7	0.2675	0.167	1.605	0.108	-0.059	0.594
ma.L8	-0.0482	0.149	-0.324	0.746	-0.339	0.243
ma.L9	0.0031	0.144	0.022	0.983	-0.278	0.284
ma.L10	0.2842	0.119	2.382	0.017	0.050	0.518
ar.S.L52	-0.7910	0.060	-13.109	0.000	-0.909	-0.673
sigma2	7.916e+08	8.68e-11	9.12e+18	0.000	7.92e+08	7.92e+08
Ljung-Box (L1) (Q):	0.06	Jarque-Bera (JB):	0.90			
Prob(Q):	0.81	Prob(JB):	0.64			
Heteroskedasticity (H):	1.16	Skew:	0.16			
Prob(H) (two-sided):	0.75	Kurtosis:	2.49			

(Better and Safer)

- Although the linear trend is observed in the seasonal decomposition, the **statistical significance** of the trend component is **not significant** (0.178)
- The p-value of 0.178 for the drift of the model when including trend can be interpreted that the trend component is **NOT statistically significant**. This means that the observed trend in the seasonal decomposition may **NOT be a real trend** but rather a result of other patterns or fluctuations in the data. So, I decided to use **trend='n'**

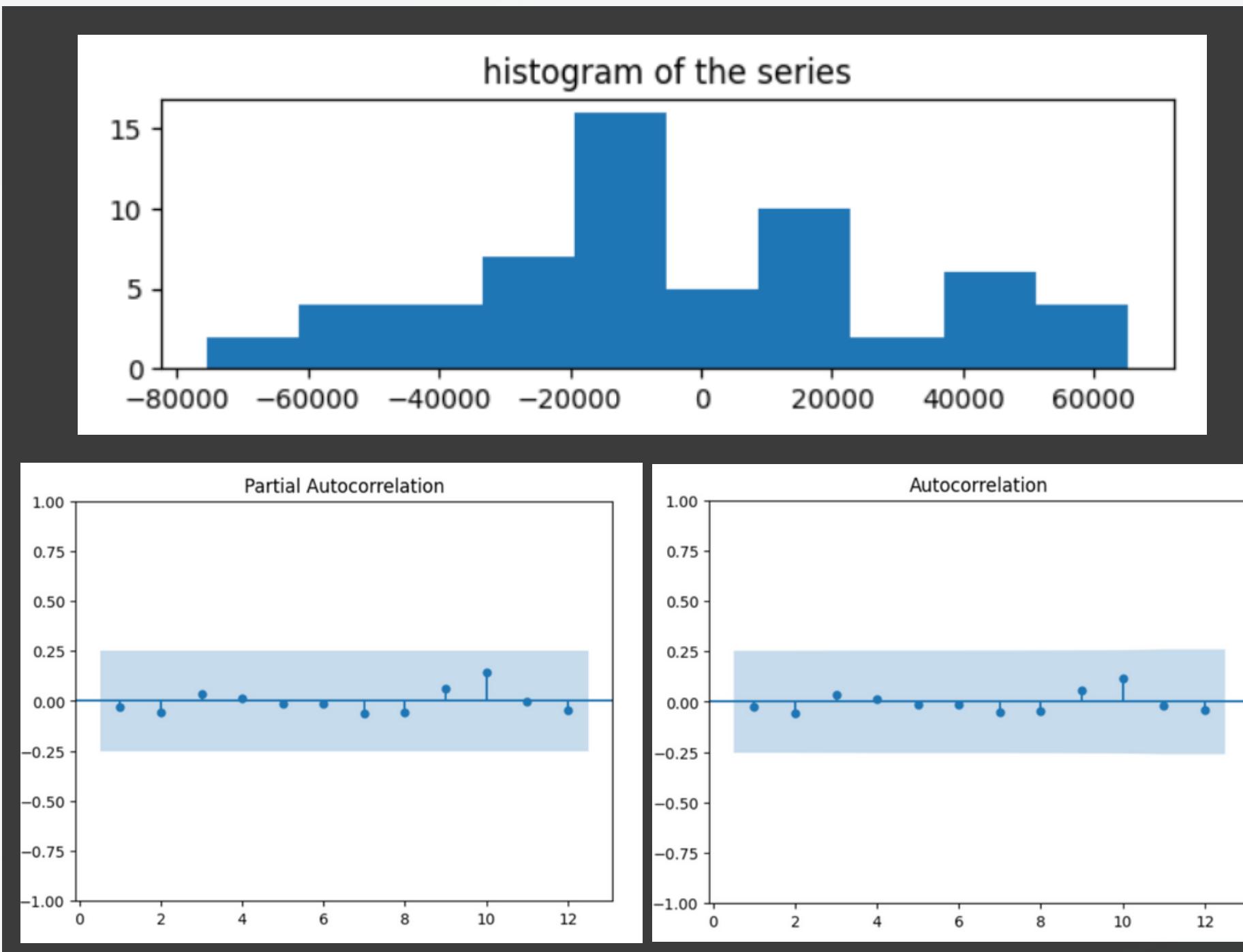


05

Modeling

BEFORE COVID MODEL

BEFORE COVID – WEEKLY RESAMPLED
(2017-08-06 TO 2019-12-29)



Check the residual of the model

- Histogram is **still not looks like normal distribution**
- PACF is in the bounds for all lags, which is now **good**
- ACF is in the bounds for all lags, which is now **good**

05

Modeling

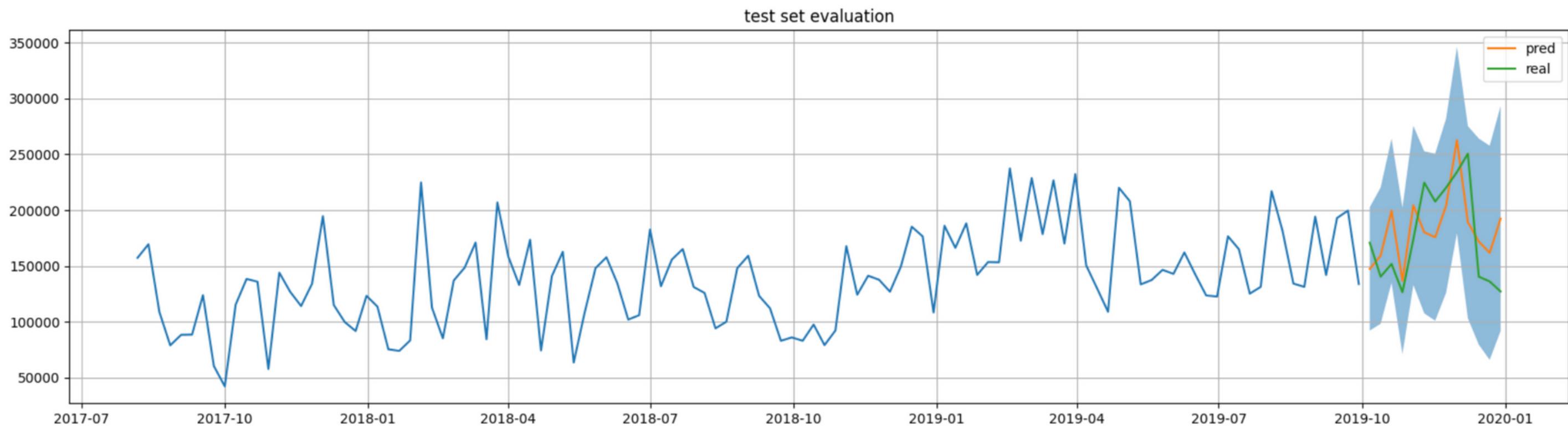
BEFORE COVID MODEL

BEFORE COVID - WEEKLY RESAMPLED
(2017-08-06 TO 2019-12-29)

MAPE

MAPE: 0.19662486408076058

Test Set Evaluation

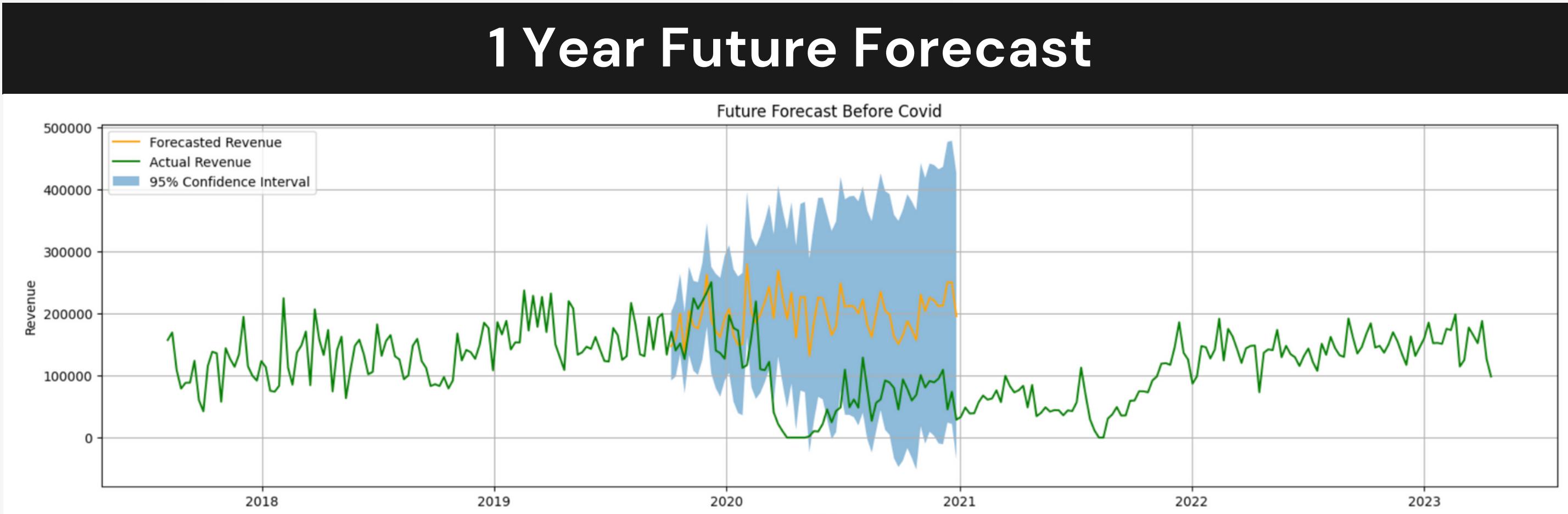


05

Modeling

FORECASTING

1 Year Future Forecast



Using this model to forecast the total revenue for **1 year** after 2019-12-29

REVENUE LOST

Loss Comparation

Date	Total_Revenue	pred_sarima	Loss
2020-03-29	10250.00	227074.902087	216824.902087
2020-04-05	0.00	191475.791397	191475.791397
2020-04-12	0.00	233730.317098	233730.317098
2020-04-19	0.00	161054.454725	161054.454725
2020-04-26	0.00	226321.549828	226321.549828
2020-05-03	0.00	226674.463260	226674.463260
2020-05-10	1951.00	132251.582862	130300.582862

Total Loss

```
total_loss = df_compare_loss['Loss'].sum()  
total_loss  
  
8426395.334673125
```

The calculated total loss in revenue that is forecasted by the model during 2020-03-29 to 2021-03-28 is about **\$8,426,395.**

05

Modeling

AFTER COVID MODEL



After Covid Data Daily-Resampled

- MAPE: 0.66



All Data Daily-Resampled

- MAPE: 0.32



All Data Weekly-Resampled

- MAPE: 0.25

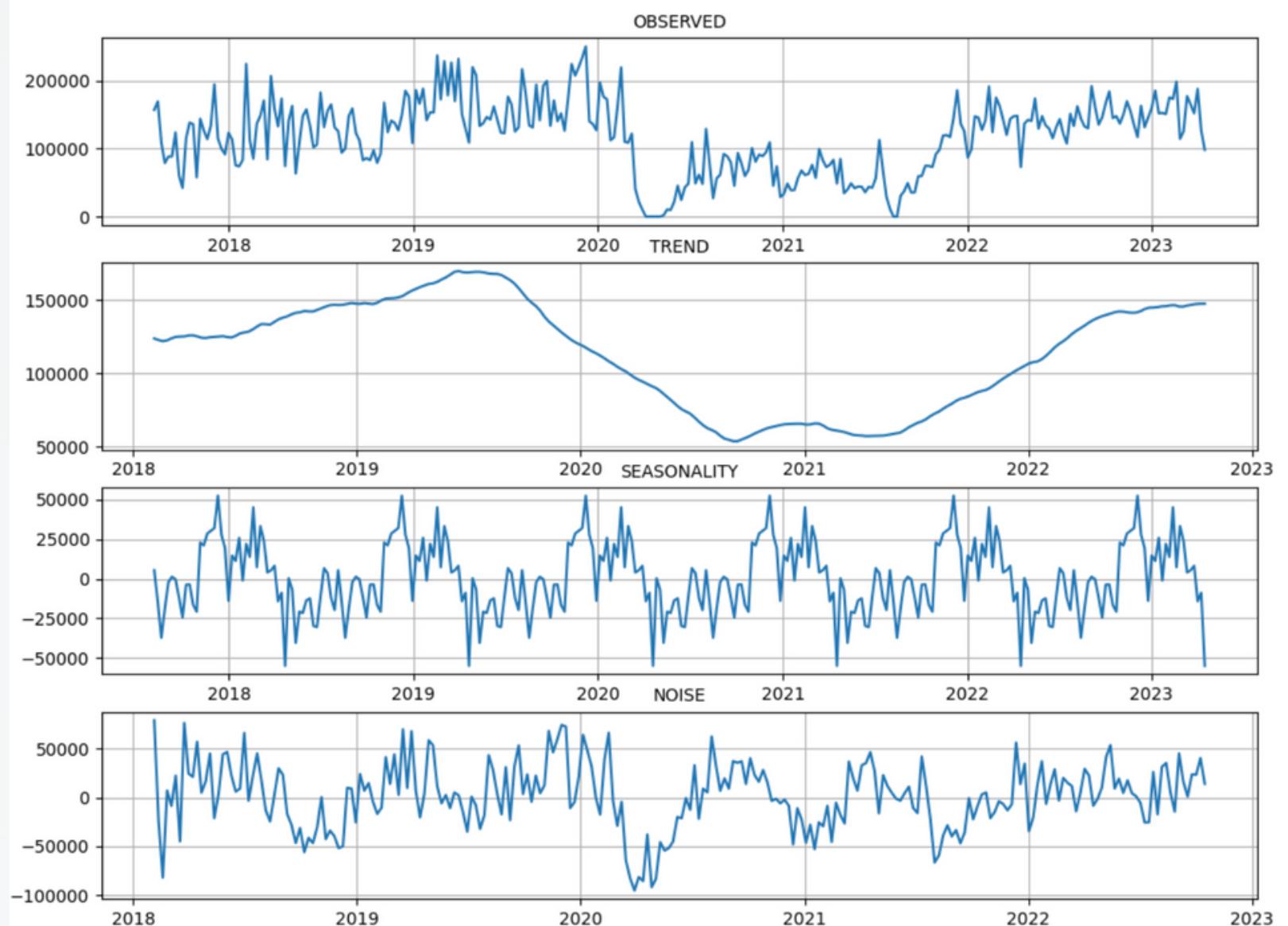
05

Modeling

AFTER COVID MODEL

ALL DATA - WEEKLY RESAMPLED
(2017-08-06 TO 2023-04-16)

seasonal_decomposed(period = 52)



Trend

No trend can be observed (Because there is a drop in the middle due to Covid-19 pandemic)

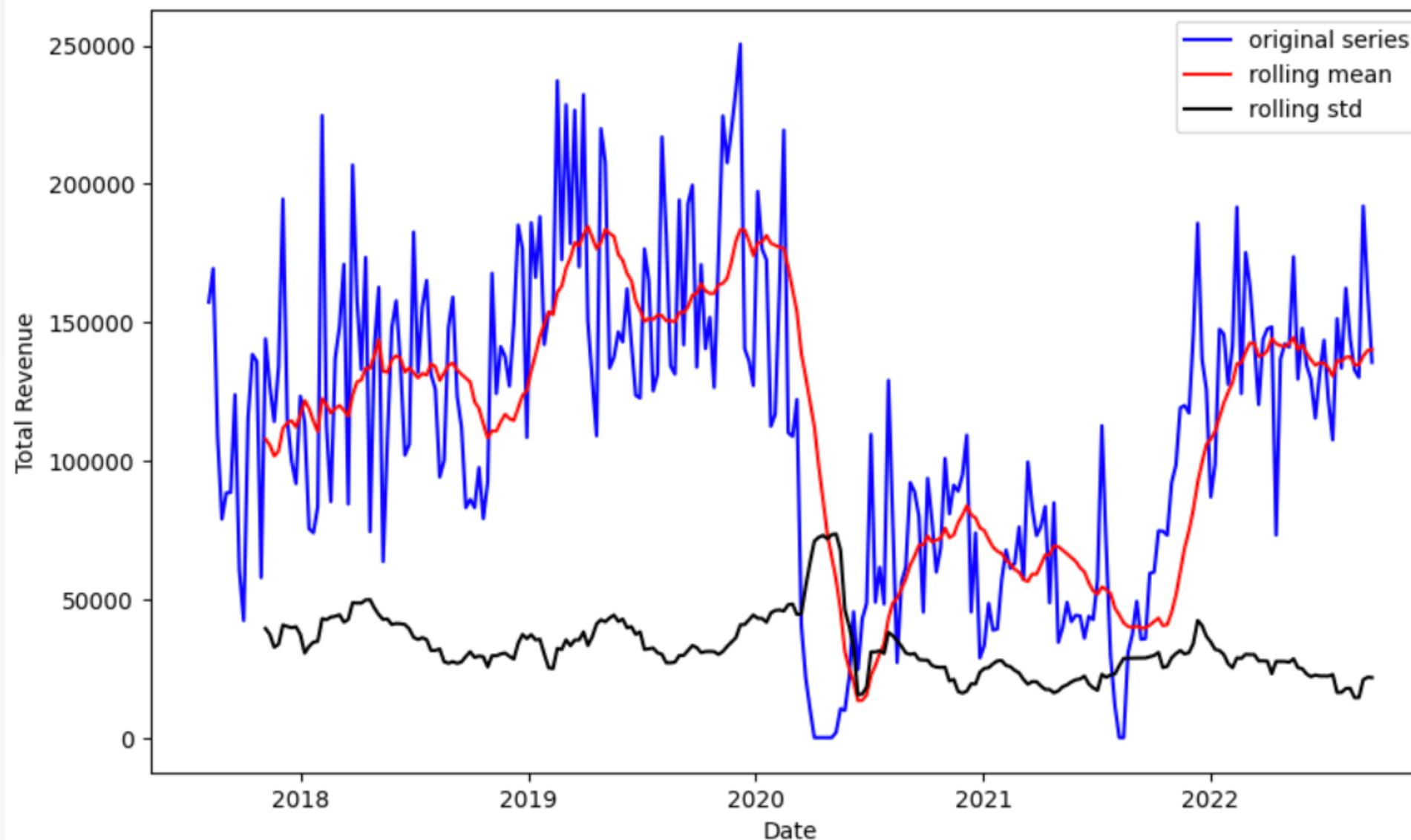
Seasonality

Seasonality can be observed

AFTER COVID MODEL

ALL DATA - WEEKLY RESAMPLED
(2017-08-06 TO 2023-04-16)

Rolling mean & std



- It can be observed that the rolling mean is **not stationary**
- From the ADF test, p-value of the data is **0.123**, which is > 0.05 that implies the **non-stationary**. So, I decide to use **SARIMA model**

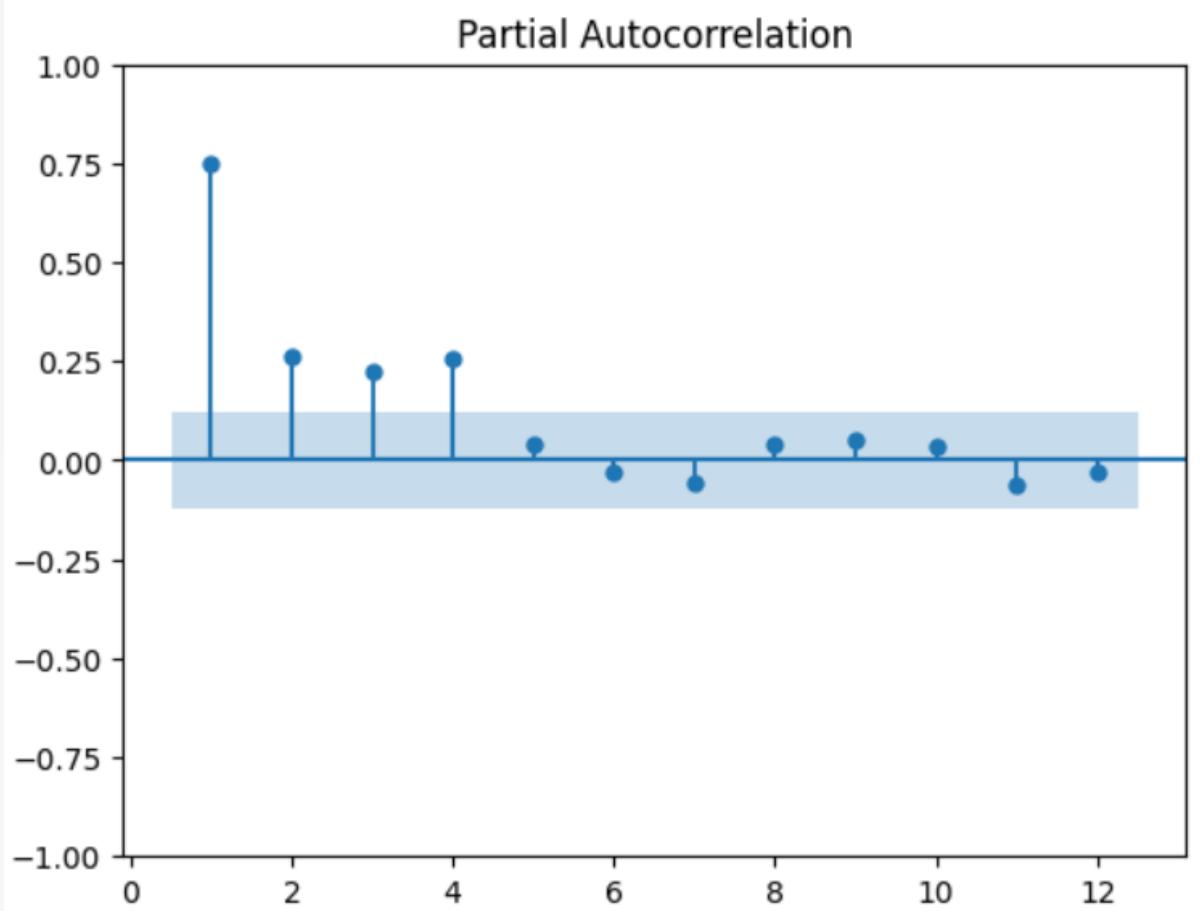
05

Modeling

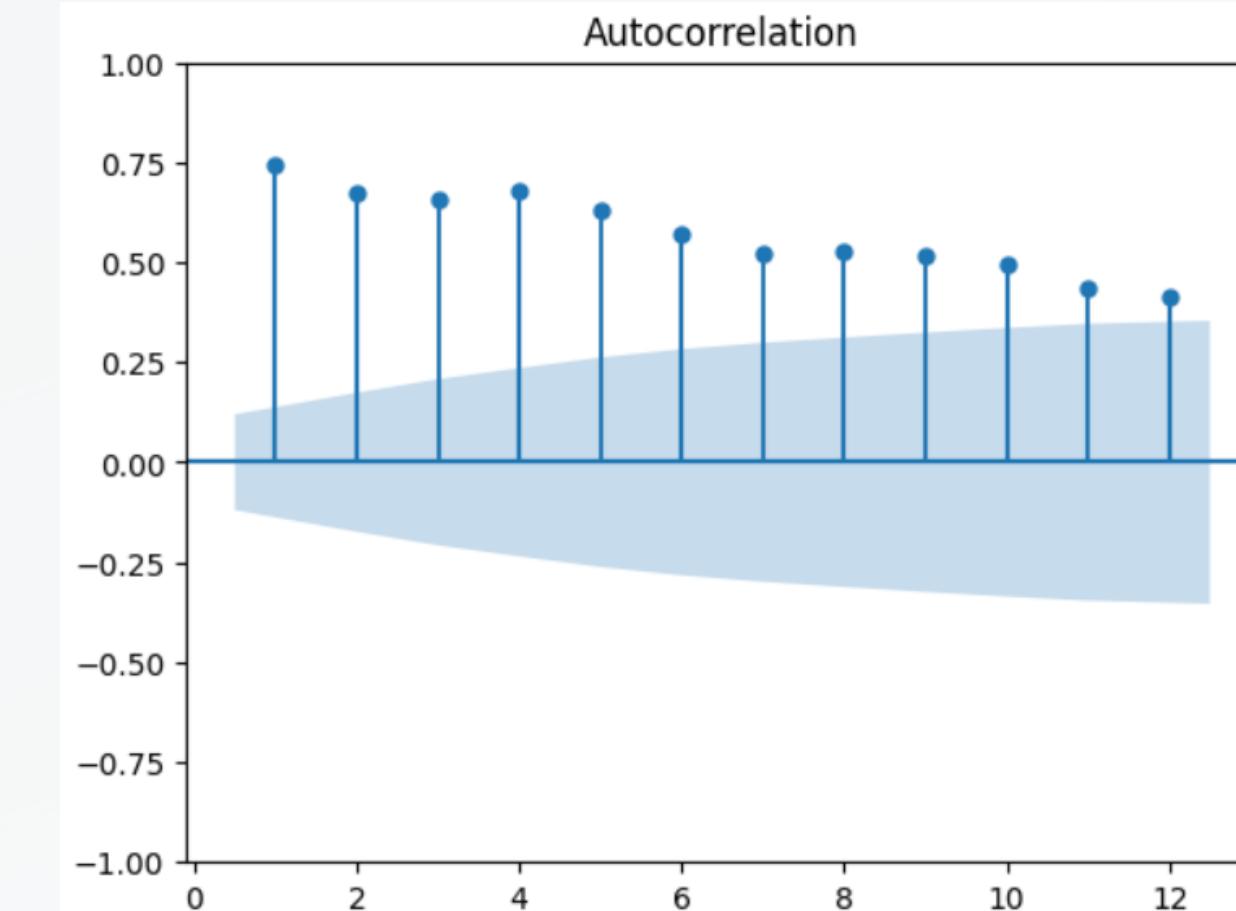
AFTER COVID MODEL

ALL DATA - WEEKLY RESAMPLED
(2017-08-06 TO 2023-04-16)

PACF



ACF



$p = 4$

$q = 0$

05

Modeling

AFTER COVID MODEL

ALL DATA - WEEKLY RESAMPLED
 (2017-08-06 TO 2023-04-16)

order = (4,1,0), seasonal_order =(1,1,0,52)

SARIMAX Results

Dep. Variable:	Total_Revenue	No. Observations:	268
Model:	SARIMAX(4, 1, 0)x(1, 1, 0, 52)	Log Likelihood	-2576.912
Date:	Thu, 27 Apr 2023	AIC	5165.825
Time:	11:27:14	BIC	5186.048
Sample:	08-06-2017 - 09-18-2022	HQIC	5173.996
Covariance Type: opg			
coef	std err	z	P> z [0.025 0.975]
ar.L1	-0.4514	0.087	-5.216 0.000 -0.621 -0.282
ar.L2	-0.3295	0.098	-3.378 0.001 -0.521 -0.138
ar.L3	-0.3064	0.088	-3.472 0.001 -0.479 -0.133
ar.L4	-0.0362	0.089	-0.406 0.685 -0.211 0.139
ar.S.L52	-0.5197	0.062	-8.373 0.000 -0.641 -0.398
sigma2	1.998e+09	1.05e-11	1.9e+20 0.000 2e+09 2e+09
Ljung-Box (L1) (Q): 0.24 Jarque-Bera (JB): 1.46			
Prob(Q): 0.63 Prob(JB): 0.48			
Heteroskedasticity (H): 0.68 Skew: -0.03			
Prob(H) (two-sided): 0.10 Kurtosis: 3.40			



order = (3,1,0), seasonal_order =(1,1,0,52)

SARIMAX Results

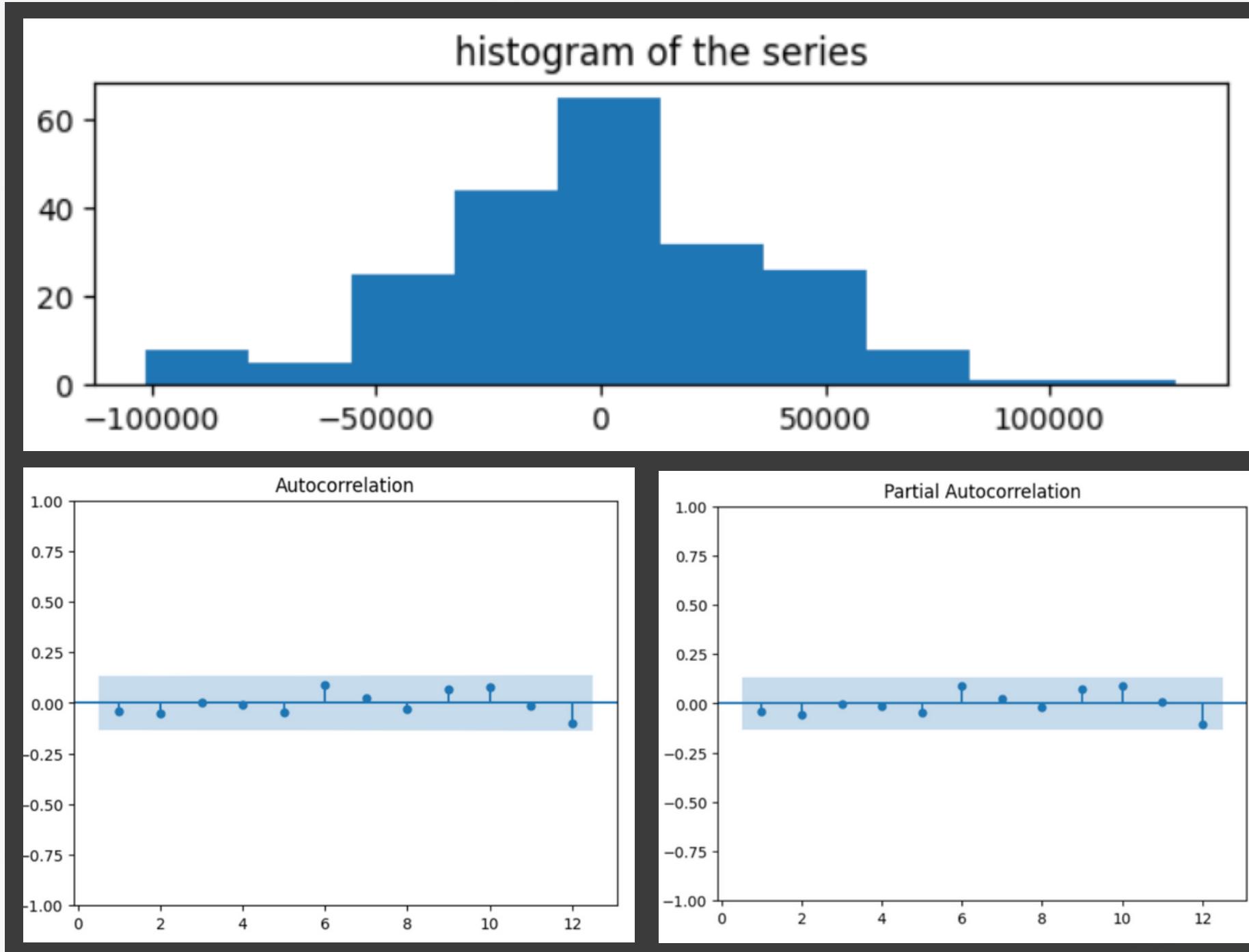
Dep. Variable:	Total_Revenue	No. Observations:	268
Model:	SARIMAX(3, 1, 0)x(1, 1, 0, 52)	Log Likelihood	-2577.220
Date:	Thu, 27 Apr 2023	AIC	5164.440
Time:	07:40:15	BIC	5181.293
Sample:	08-06-2017 - 09-18-2022	HQIC	5171.250
Covariance Type: opg			
coef	std err	z	P> z [0.025 0.975]
ar.L1	-0.4406	0.083	-5.302 0.000 -0.604 -0.278
ar.L2	-0.3181	0.095	-3.365 0.001 -0.503 -0.133
ar.L3	-0.2912	0.080	-3.644 0.000 -0.448 -0.135
ar.S.L52	-0.5212	0.062	-8.443 0.000 -0.642 -0.400
sigma2	2.011e+09	1.77e-12	1.14e+21 0.000 2.01e+09 2.01e+09
Ljung-Box (L1) (Q): 0.35 Jarque-Bera (JB): 1.14			
Prob(Q): 0.55 Prob(JB): 0.56			
Heteroskedasticity (H): 0.68 Skew: -0.02			
Prob(H) (two-sided): 0.10 Kurtosis: 3.35			



- In can be observed that the order of (4,1,0) from the previous slide results in getting the p-value of ar.L4 = **0.685**, which is > 0.05. So, the solution is to stepping the order down to (3,1,0) and check the p-value again, which ar.L3 is < 0.05, which is now **significant**
- In this case, trend will be set to None since there is **no trend** observed from the seasonal_ decompose

AFTER COVID MODEL

ALL DATA - WEEKLY RESAMPLED
(2017-08-06 TO 2023-04-16)



Check the residual of the model

- Histogram is **normal distribution**, which is **good**
- PACF is in the bounds for all lags, which is **good**
- ACF is in the bounds for all lags, which is **good**

05

Modeling

AFTER COVID MODEL

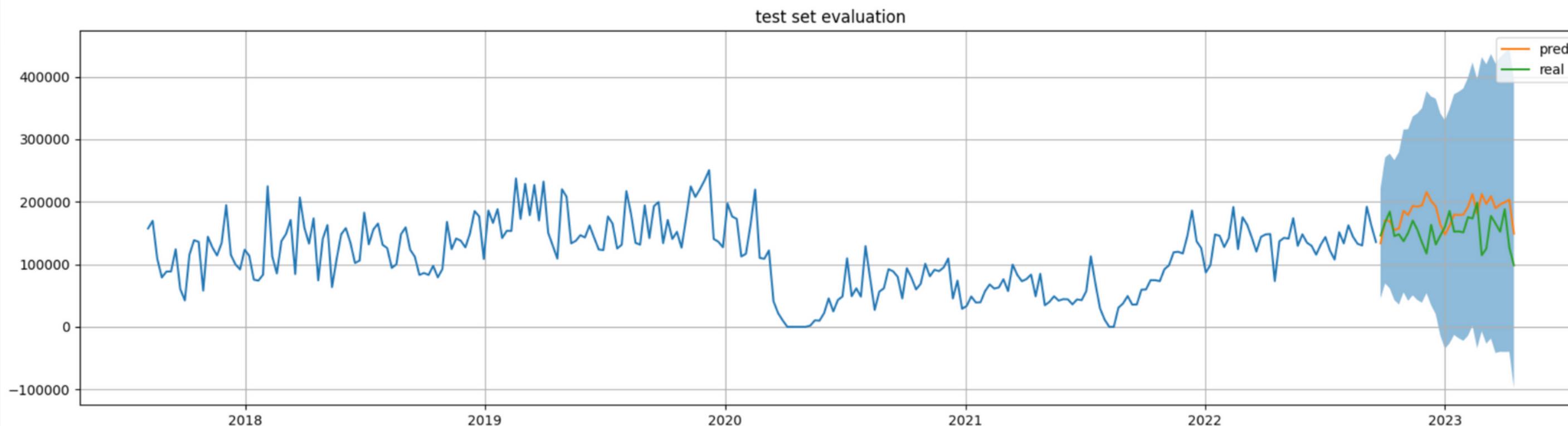
ALL DATA - WEEKLY RESAMPLED
(2017-08-06 TO 2023-04-16)

MAPE

MAPE: 0.25663129769823656

The error are mainly came from the **covid pandemic** causing the **huge drop** on the of total revenue for about whole 1.5 years, which may affect the accuracy of the model

Test Set Evaluation

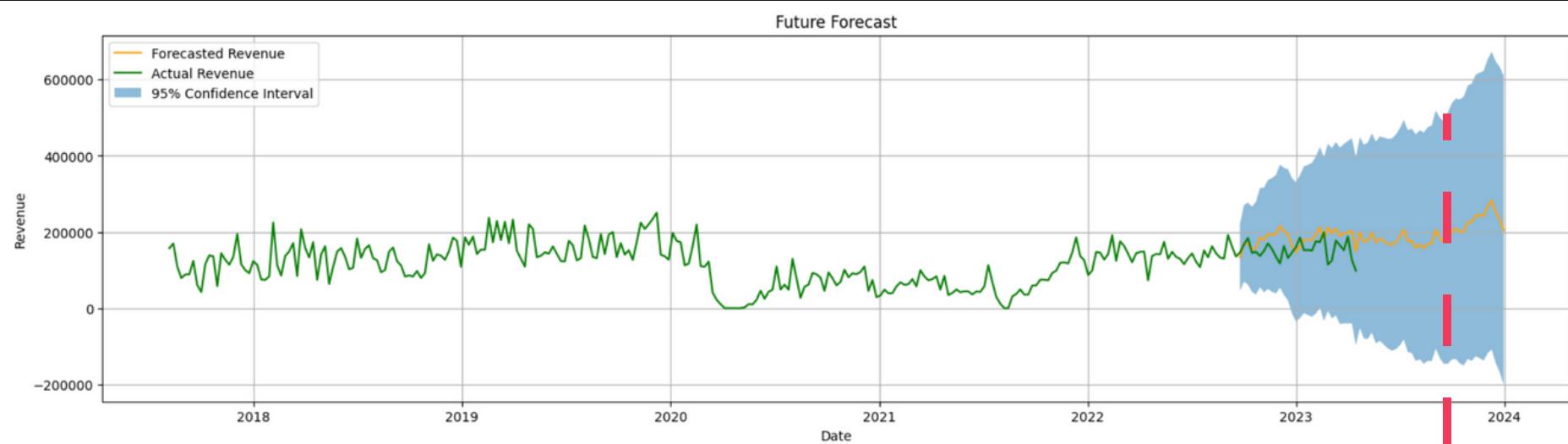


05

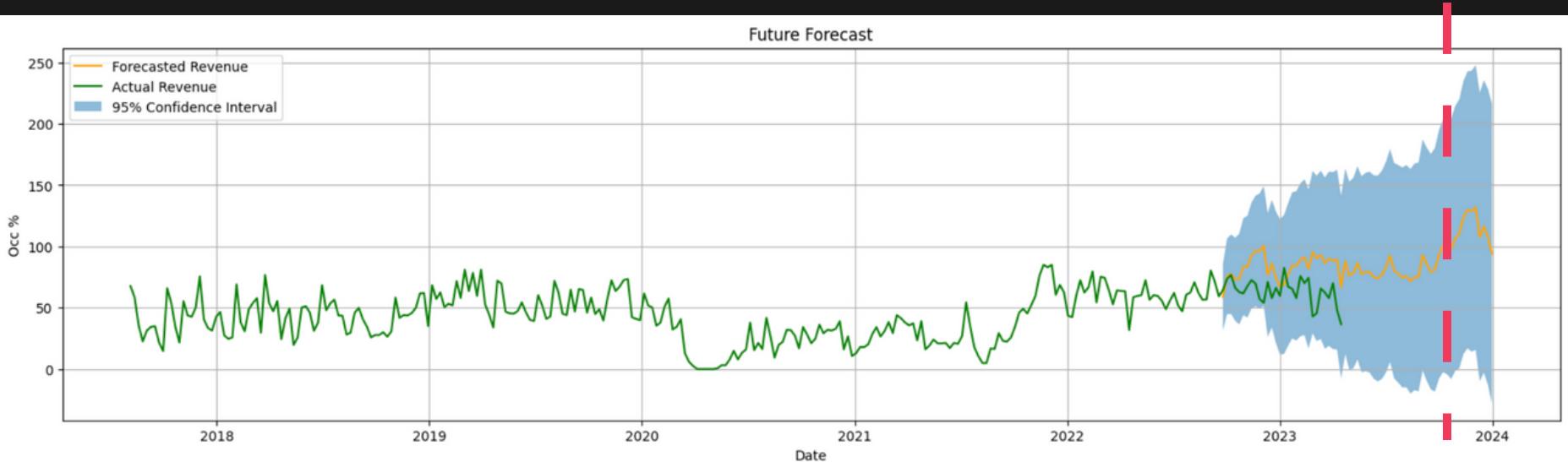
Modeling

FORECASTING

2023 Total Revenue Model Forecast



2023 Occ % Model Forecast



2023-10-08

- We also have to check that whether the **Occ %** is **exceeding 100%** or not

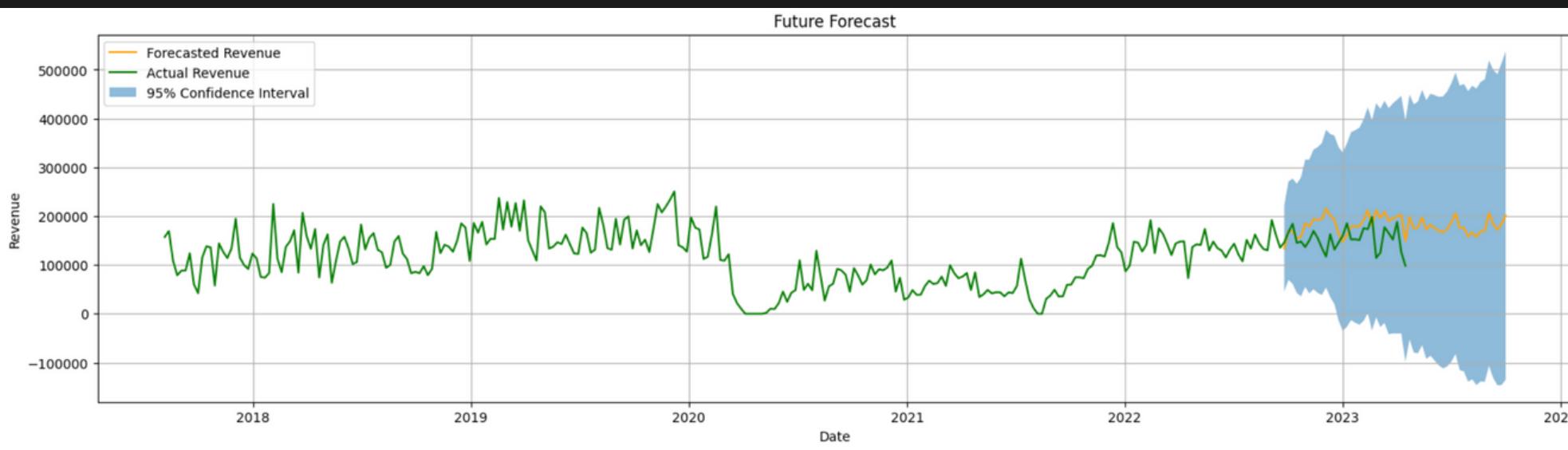
- It can be observed that the percentage of **occupancy rate exceed 100%** since **2023-10-08**, which implies that our model **CANNOT** be used to forecast further than that day.

05

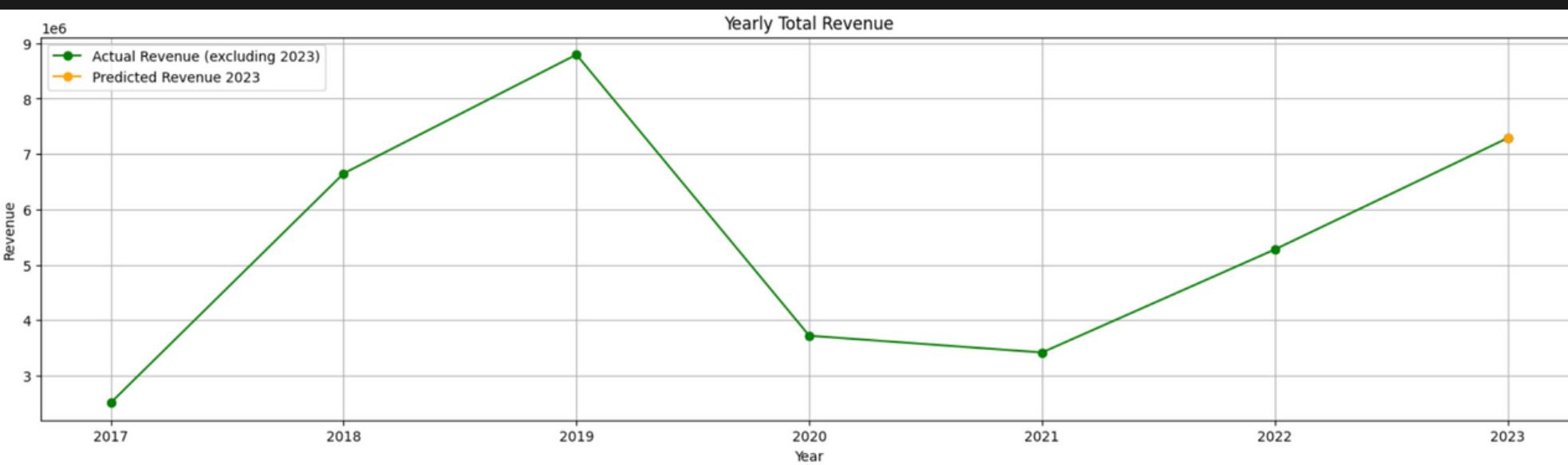
Modeling

FORECASTING

2023 Total Revenue Forecast until 2023-10-08



Yearly Total Revenue + 2023 Forecasted



Total_Revenue

Date

2017	2.509827e+06
2018	6.649369e+06
2019	8.794506e+06
2020	3.719671e+06
2021	3.417288e+06
2022	5.277910e+06
2023	7.288862e+06

05

Modeling

FORECASTING

2023 Total Revenue Forecast until 2023-10-01 + Last Quarter



Date	Total_Revenue
2017	2.509827e+06
2018	6.649369e+06
2019	8.794506e+06
2020	3.719671e+06
2021	3.417288e+06
2022	5.277910e+06
2023	9.718483e+06

Since our model does the forecast until 2023-10-01. So, in order to find the total revenue on 2023 we have to do consider the **last Quarter** by multiplying the total revenue forecasted by the model, which is \$7,288,862, by 4/3. So, our final total revenue will be **\$9,718,483**.

06

Result Analysis

RESULT ANALYSIS & SUGGESTED BUSINESS ACTION

Caution

- This is the suggestion from interpreting EDA and model, which still have an **error**

Result Analysis

- Based on the current model, the revenue **loss** from Covid-19 pandemic is about **\$8,426,395**
- The predicted total revenue in 2023 is about **\$9,718,483**, which is now on an **increasing trend** and also **higher than** before Covid-19 pandemic
- Based on the occupancy percentage model, it is projected that occupancy rates will surpass 100% by **October 8th, 2023**. As a result, **marketing efforts** can be intensified to maximize revenue until that date. Beyond that point, the hotel may need to consider **raising room prices**, as occupancy levels will be at capacity.

Suggested Business Action

- The hotel should increase the capability of satisfying business people by:
 - ✓ Increasing the speed of the **Wi-Fi**
 - ✓ Expanding the **working space** for customers
 - ✓ Enhance the **pick-up facilities** for customers that come from airport
 - ✓ Invest on promoting with **Google platform** to increase the chance of customer finding the hotel by searching with keywords
 - ✓ Invest on **OTA** to drive & increase total revenue

THANK YOU

EDA Code: <https://colab.research.google.com/drive/1ohBuqJdljXI64h99PUGkESCGYWnmh6RU?usp=sharing>.

Modeling Code: https://colab.research.google.com/drive/1VA1PQ_FDwKUFQE86mEuwg3wWKddNwv?usp=sharing.