

Final Project

Data Science & Machine Learning

Bootcamp - Remote

Tiago Santos

2025



Project

Data - Reservation report from an Apartment Hotel

- **EDA - Use to analyse the house keeping services in a Hotel**
- **Create a ML model to predict HK services**
 - **Create a dashboard in tableaux**
 - **App for guest information's**

Data Cleaning

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19463 entries, 0 to 19462
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Lista de reservas  19463 non-null  object
1   Unnamed: 1          19463 non-null  object
2   Unnamed: 2          19463 non-null  object
3   Unnamed: 3          19463 non-null  object
4   Unnamed: 4          19463 non-null  object
5   Unnamed: 5          19463 non-null  object
6   Unnamed: 6          19463 non-null  object
7   Unnamed: 7          19463 non-null  object
8   Unnamed: 8          19463 non-null  object
9   Unnamed: 9          16903 non-null  object
10  Unnamed: 10         19463 non-null  object
11  Unnamed: 11         19463 non-null  object
12  Unnamed: 12         17688 non-null  object
13  Unnamed: 13         19463 non-null  object
14  Unnamed: 14         17635 non-null  object
15  Unnamed: 15         16074 non-null  object
16  Unnamed: 16         16361 non-null  object
17  Unnamed: 17         15870 non-null  object
18  Unnamed: 18         984 non-null    object
19  Unnamed: 19         2151 non-null   object
dtypes: object(20)
memory usage: 3.0+ MB
```

```
resv_id          object
resv_date        datetime64[ns]
resv_status      object
checkin          datetime64[ns]
checkout         datetime64[ns]
rn              Int64
resv_total       object
adults           Int64
child            Int64
baby             Int64
room_type        object
id_roomtype      Int64
id_client        Int64
guest            object
client_country   object
guest_country    object
guest_language   object
distribution     object
dtype: object
```

creating new columns names

Converting data types,

Clean empty spaces

Column standardization

Data Cleaning

resv_status		oos	
Paga	9272	0	10856
Cancelada	7206	m	539
De proprietário	1341	h	344
Pré-reserva	966	a	48
Indisponível	402	t	46
Confirmada	250	ADM	2
Pedido informação	12	Overbooking	1
Sob pedido	10	maintenance	1
Pedido de disponibilidade	2	team	1
Garantia	1	hk	1
Name: count, dtype: int64		Name: count, dtype: int64	

resv_status		resv_id	0
Paga	9272	resv_date	0
De proprietário	1341	resv_status	0
Pré-reserva	966	checkin	0
Confirmada	250	checkout	0
Sob pedido	10	rn	0
Name: count, dtype: int64		resv_total	0
		adults	0
		child	0
		baby	90
		room_type	0
		id_roomtype	0
		id_client	0
		guest	0
		client_country	67
		guest_country	227
		guest_language	0
		distribution	0
		dtype: int64	

resv_status		resv_id	0
Active	10498	resv_date	0
Out_of_Service	1341	resv_status	0
Name: count, dtype: int64		checkin	0
		checkout	0
		rn	0
		resv_total	0
		adults	0
		child	0
		baby	90
		room_type	0
		id_roomtype	0
		id_client	0
		guest	0
		client_country	67
		guest_country	227
		guest_language	0
		distribution	0
		dtype: int64	

***Replacing some value counts**

***Drop some rows Maintenance, cancelation, request's reservations/ block's**

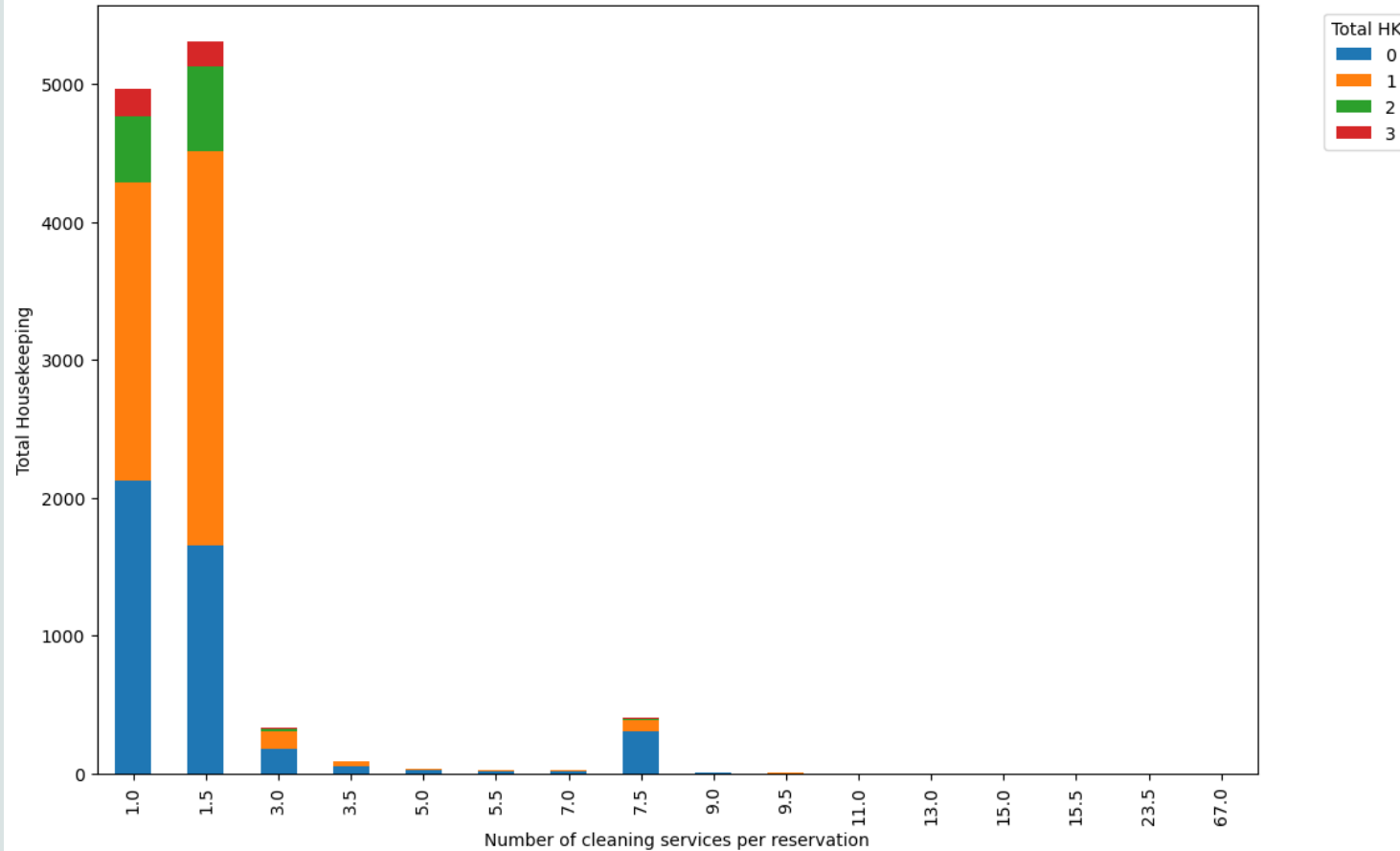
*** Drop some categorical columns**

- Refiled baby, guest country
- Mapping - Room Types, Resv Satus,
 - Changing checkout date
 - Copy
- **Generate new features - Hk full stay** (end of reservation) **and Hk half stay** (after 3 day's), **Total HK** (sum of hk full stay and hk half stay)

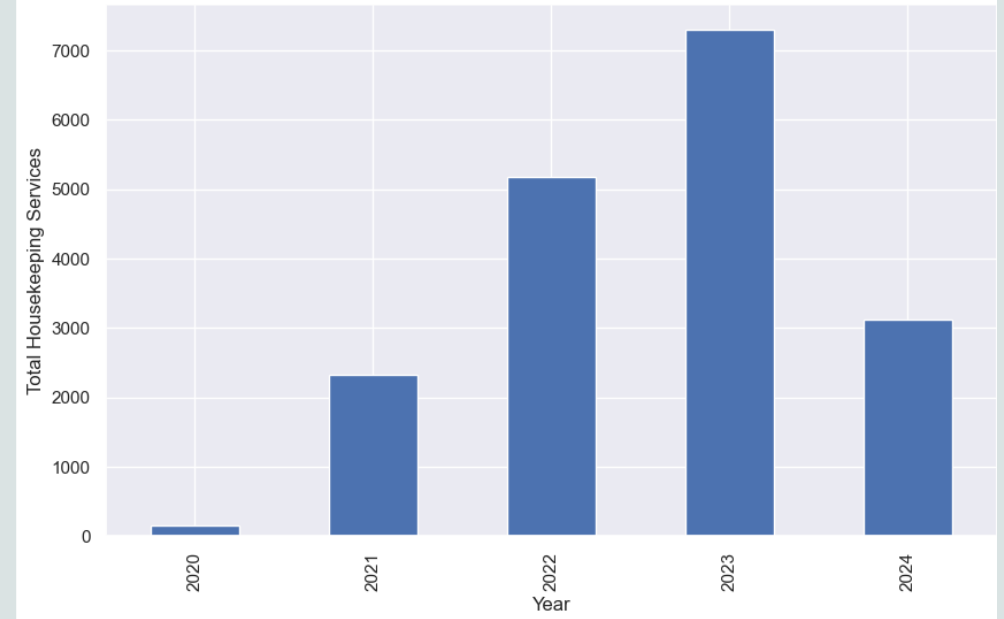
	resv_status	rn	adults	child	baby	room_type	id_client	oos	hk_Full_stay	checkout_year	checkout_month	checkout_day	checkout_day_name	hk_half_stay	total_HK
6416	1	7	1	0	0	2	6299235	1	1	2021	8	7	7	0.5	1.5
886	1	3	2	0	0	0	14452319	1	1	2023	4	23	1	0.0	1.0
152	1	7	2	0	0	0	13166238	1	1	2023	6	24	7	0.5	1.5
10249	1	7	1	0	0	2	6293826	1	1	2023	8	13	1	0.5	1.5

EDA

Crosstab of Room Type and Total Housekeeping



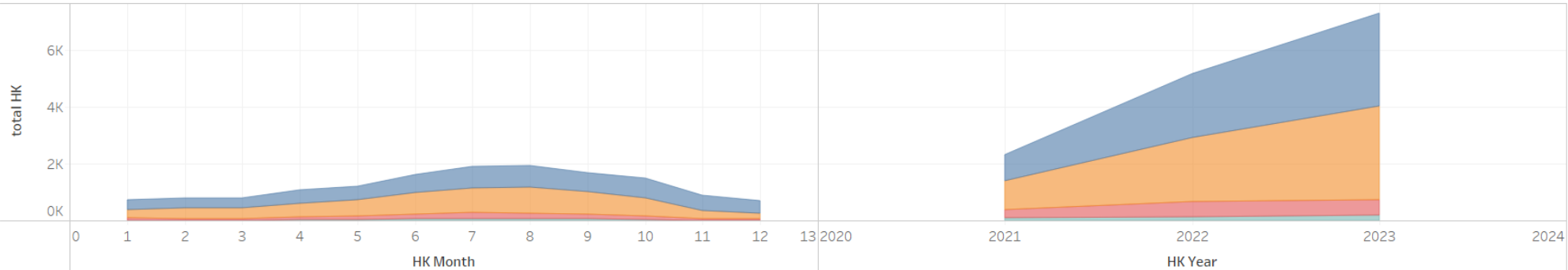
Total Housekeeping Services by Year



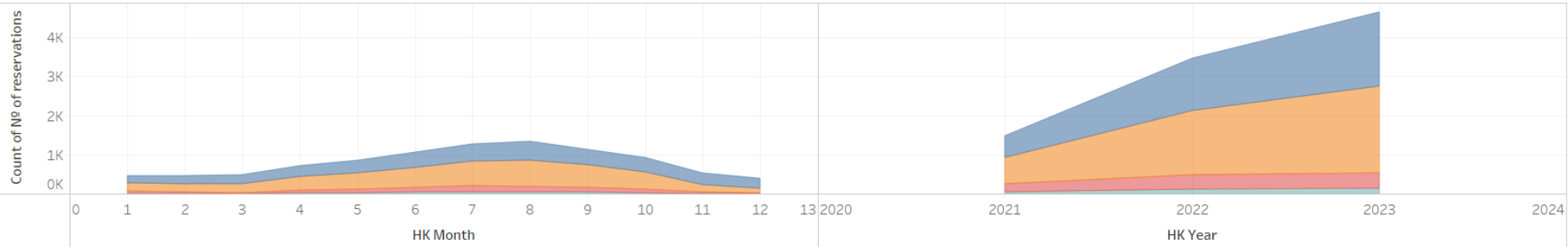
EDA - tableau

HK vs sales

Hk Services



Sales - Reservations made



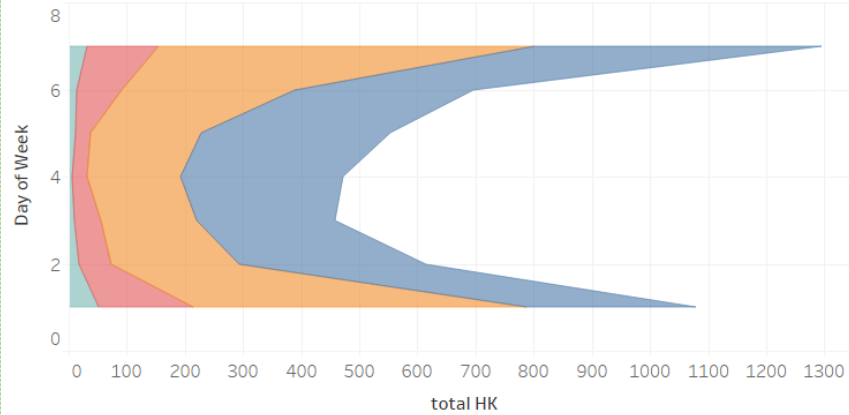
Hk Services

Roo..	HK Month / HK Year																													
	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023	2021	2022	2023
1	5.5	55.5	211.5	9.5	40.5	313.5	9.5	53.0	318.5	32.0	131.5	299.0	68.0	198.0	303.0	121.5	355.5	289.5	168.5	369.5	341.5	254.5	342.5	337.0	183.0	325.0	284.5	9.5	40.5	313.5
0	11.0	61.0	271.5	9.5	63.0	269.0	14.0	84.0	267.5	23.5	146.0	299.5	43.0	166.5	277.0	122.5	255.0	271.5	165.0	289.5	291.0	182.0	279.0	290.0	156.5	237.5	268.5	13.0	58.0	277.0
2	4.5	32.0	52.5	16.0	13.5	29.0	1.5	18.5	22.5	22.5	29.5	45.0	20.5	58.0	47.5	45.0	66.0	48.0	41.0	90.0	71.5	48.0	68.0	63.0	41.0	71.5	50.0	3.0	13.5	29.0
3		1.0	10.5		2.0	13.5		1.0	14.5	5.0	8.0	30.0	10.5	12.0	17.0	9.5	22.5	23.5	17.0	29.0	23.0	25.0	25.5	23.0	12.0	21.0	21.0		10.5	30.0

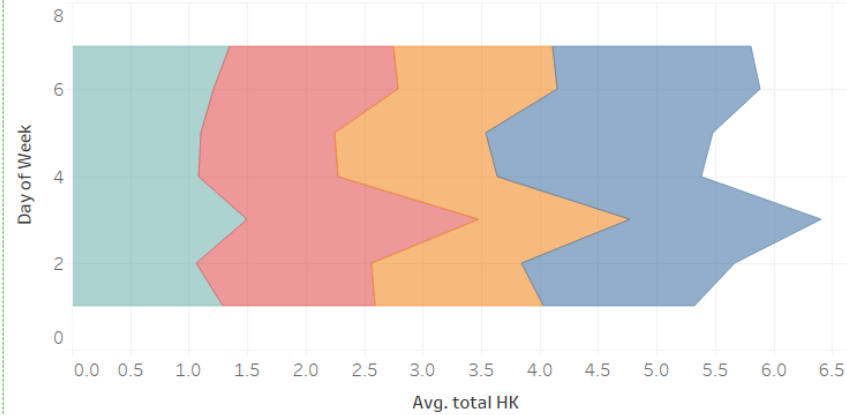
EDA - tableau

HK Services vs LYSD

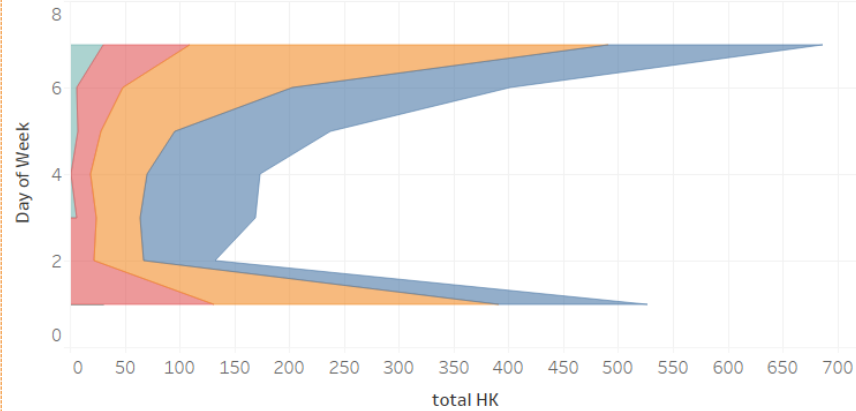
Hk Services - Count



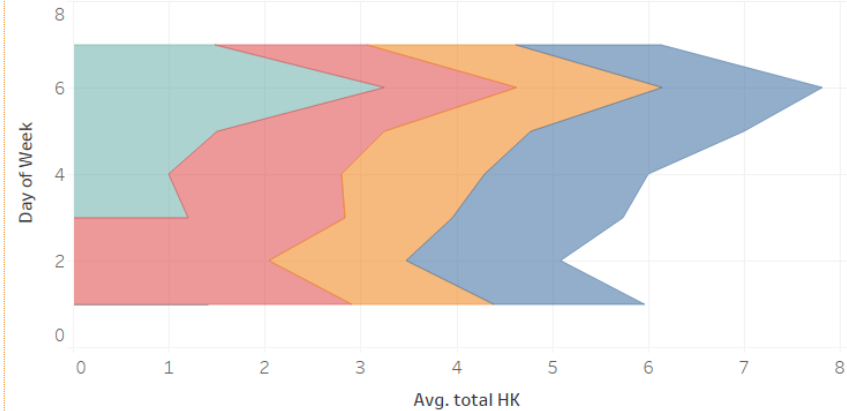
Hk Services - Average



Hk services * LYSD - Count



Hk services * LYSD - Average



Room Type



Resv Status



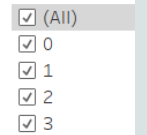
HK Month



Year



Room Type

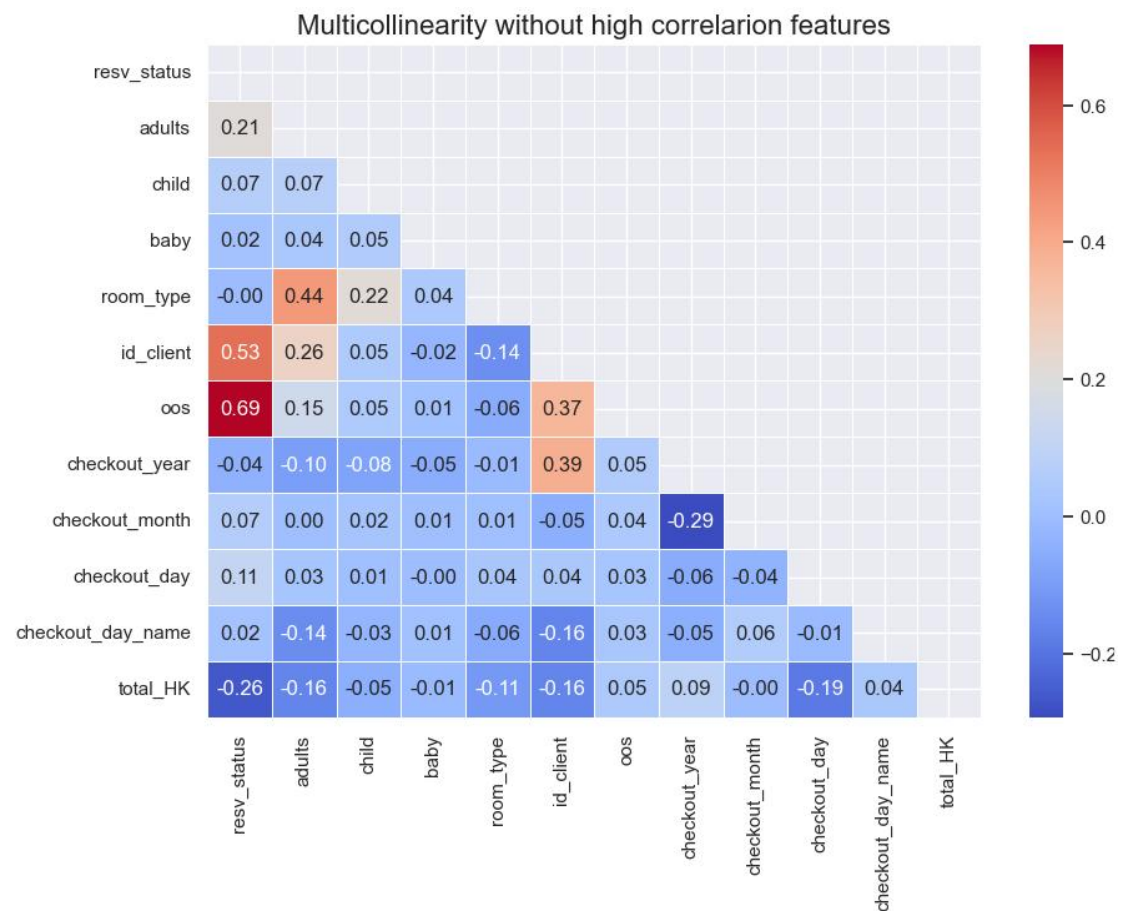


Dif Year



Correlation & multicollinearity

```
total_HK      1.000000
hk_half_stay  0.988149
hk_Full_stay  0.986008
rn            0.971654
checkout_year 0.088207
oos           0.048232
checkout_day_name 0.036854
checkout_month -0.002626
baby         -0.010318
child        -0.050501
room_type    -0.110259
id_client    -0.159585
adults       -0.161237
checkout_day -0.187876
resv_status  -0.264344
Name: total_HK, dtype: float64
```



Models

100% of our data: 11201.
70% for training data: 7840.
30% for test data: 3361.

	actual	pred	diference
12212	1.5	1.508693	0.008693
2575	1.5	1.453397	0.046603
2344	1.0	1.338038	0.338038
4686	1.0	1.381479	0.381479
4806	1.0	1.423327	0.423327
12372	1.5	1.500530	0.00053
7848	1.0	1.322118	0.322118
482	1.0	1.287037	0.287037

R2 = 0.1538
RMSE = 1.65
MSE = 2.73
MAE = 0.64

Linear regression

R2 = 0.2548
RMSE = 1.5516
MSE = 2.4075
MAE = 0.3831

Decision tree

R2 = 0.3466
RMSE = 1.4529
MSE = 2.1109
MAE = 0.4127

Xgboost - the best model

Data & future improvements

Create new features – generating new KPI's :

- **Average clean per room type**


- **Time per cleaning** By defining av. time cleaning by room types
- **How many team members** is needed for each day
 - Try news models, hipertuning...


** form this kind of approach I've secured many jobs positions and also create new ones**


chatbot

It's Working! 😊

Orange Hotel Service Chatbot (RAG)


 can you tell me the name of the hotel

 I'm sorry, the name of the hotel isn't mentioned in the information provided. Let me check with our team member.

 can you tell me the services hours

 Sure, here are our service hours:

- Check-in: 2:00 PM to 11:00 PM
- Check-out: until 12:00 PM
- Breakfast: 7 AM to 10:30 AM
- Lunch: 12:00 PM to 2:00 PM

 what kind of services do you have

 We offer a variety of services including a spa, free wi-fi, fitness centre, swimming pool, restaurant, and transfer services.

Ask about our hotel services, or type your reservation ID...



chatbot

```
hotel_info.md > # Hotel Orange > ## Hotel Service Hours
1  # Hotel Orange
2  ## 4 star Apartment Hotel ( **** )
3
4  ## Hotel Service Hours
5  - Check-in: 2:00 PM to 11:00 PM
6  - Check-out: until 12:00 PM
7  - Breakfast: 7 AM to 10:30 AM
8  - Lunch: 12:00 PM to 2:00 PM
9  - Dinner: 6:00 PM to 10:00 PM
10 - Spa and Wellness centre: 9:00 AM to 8:00 PM
11
12 ## Amenities
13 We have a spa, free wi-fi, fitness centre, swimming pool, restaurant, transfer services.
14
15 ## Reservations Meal Plan
16 - all Hotel reservations has included Bed and Breakfast (meal Plan)
17
18 ## Reservation Cancellation Policy
19 - free of charge if the cancellation is made 3 days before of arrival time
20
21 ## Late check-out
22 - Available on request on the day of check-out, may charges be applied
23
24 ## Early check-in
25 - Available on request on the day of check-in, may charges be applied
26
27 ## Hotel Pet Policy
28 - On request (additional charge), please contact the reservations department at reservations@hotelorange.com
29
30 ## Room/Apartment Types
31 - Studio
32 - One Bedroom Apartment
33 - Two Bedroom Apartment
34 - Three Bedroom Apartment
35
36 ## Apartments views
37 - all the apartments has pool view
38
39 ## Maximum occupancy by room/apartment type
40 - Studio (maximum occupancy 2 persons)
41 - One Bedroom Apartment (maximum occupancy 3 persons)
42 - Two Bedroom Apartment (maximum occupancy 5persons)
43 - Three Bedroom Apartment (maximum occupancy 7 persons)
```

- Load my data set – Clean . csv
- Load my hotel information. Md file
 - Split into chunks
- Embeddings vectors with OpenAI and Chroma DB
- Function to look for the reservation
 - Create the Streamlit app

Thank you!