Part 1 - METHODOLOGY DOCUMENT

AIR BNB NYC CASE STUDY

Analysis Details

ABSTRACT

This Document Details the Methodology of Presentations 1

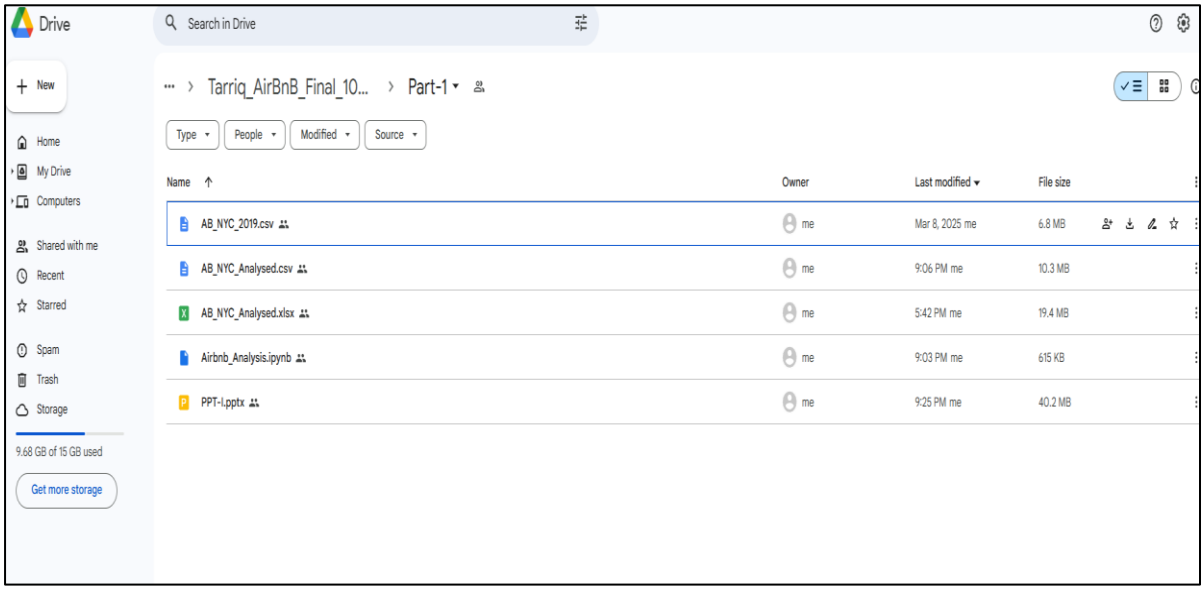
Tarriq Ferrose Khan

May 2024 DSC 68

Contents

Files Created to Prepare the Presentations (Jupyter notebook, PPT, etc)

G Drive Folder Link - All Files Developed For Presentation - 1



PPT 1

Methodology - EDA

- ✓ A notebook file is developed for EDA and Data Modelling
- ✓ User Defined Functions: Several User Defined functions developed to analyse the data for:
- ✓ Null Columns,
- ✓ Columns By Type
- ✓ Create Bin Values
- ✓ Draw Charts and other purposes.

- ✓ Null Value Handled and replaced with respective values.

- ✓ Outliers Checked through boxplots and removed appropriately.
- ✓ Date columns formatted.

Used Create Bins Method and created Group Column for Analysis.
Created Additional Columns on the Data Frame for Furnished etc.,
Implemented the Text Processing using Regex to derive Bedroom Spec and other specification columns

Data Modelling

Used Linear Regression to create a model with Price as the dependent Variable to understand the Price Impacting features.

OLS Regression Results

Dep. Variable:	price	R-squared:	0.833
Model:	OLS	Adj. R-squared:	0.833
Method:	Least Squares	F-statistic:	1.046e+04
Date:	Sun, 09 Mar 2025	Prob (F-statistic):	0.00
Time:	18:55:57	Log-Likelihood:	93415.
No. Observations:	33656	AIC:	-1.868e+05
Df Residuals:	33639	BIC:	-1.867e+05
Df Model:	16		
Covariance Type:	nonrobust		

```
from sklearn.metrics import r2_score
r2_score(y_test_new, y_pred_cnt_new)
```

0.0s

```
0.8059996807702011
```

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0158	0.000	34.547	0.000	0.015	0.017
price_300-600	0.0511	0.000	129.815	0.000	0.050	0.052
price_600-900	0.1169	0.001	130.122	0.000	0.115	0.119
price_900-1200	0.1741	0.001	117.794	0.000	0.171	0.177
price_Greater Than 1200	0.4225	0.001	311.872	0.000	0.420	0.425
Private room	-0.0142	0.000	-82.674	0.000	-0.015	-0.014
Shared room	-0.0185	0.001	-33.929	0.000	-0.020	-0.017
Brooklyn	0.0123	0.000	26.376	0.000	0.011	0.013
Manhattan	0.0178	0.000	38.261	0.000	0.017	0.019
Queens	0.0100	0.001	19.697	0.000	0.009	0.011
Staten Island	0.0090	0.001	8.686	0.000	0.007	0.011
AVL_365_100-200	0.0021	0.000	7.823	0.000	0.002	0.003
AVL_365_200-300	0.0020	0.000	7.064	0.000	0.001	0.003
AVL_365_300-400	0.0037	0.000	15.378	0.000	0.003	0.004
RVWSPERMONTH_Greater Than 8	-0.0031	0.002	-2.048	0.041	-0.006	-0.000
NUM_RVW_60-120	-0.0007	0.000	-2.299	0.022	-0.001	-0.000
NUM_RVW_Greater Than 120	-0.0012	0.000	-2.842	0.004	-0.002	-0.000

Equation from the predicted Model

const x 0.0158 + price_300-600 x 0.0511 + price_600-900 x 0.1169 + price_900-1200 x 0.1741 + price_Greater Than 1200 x 0.4225 + Private room x -0.0142 + Shared room x -0.0185 + Brooklyn x

Charts

Prepared using Excel Pivot and placed in PPT as appropriate

-x-End of Document-x-