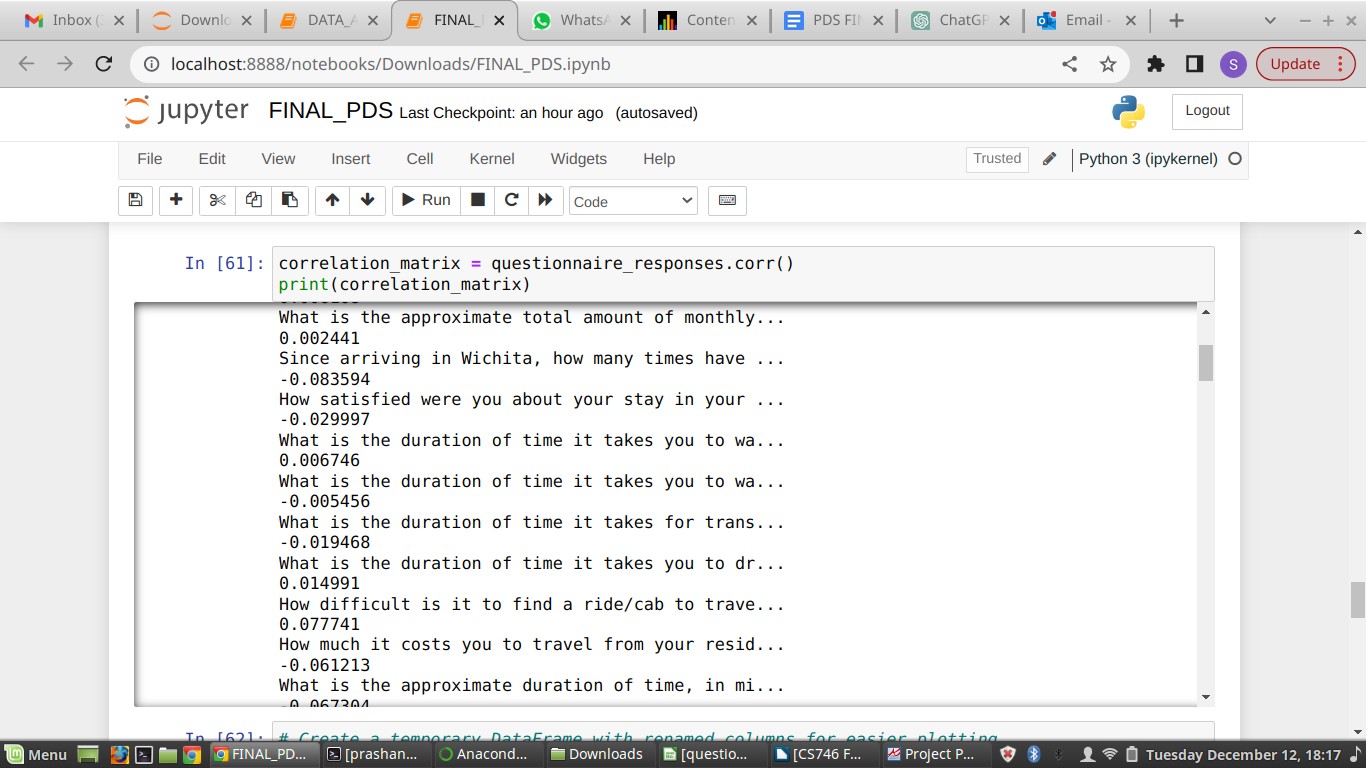
# PDS FINAL REPORT

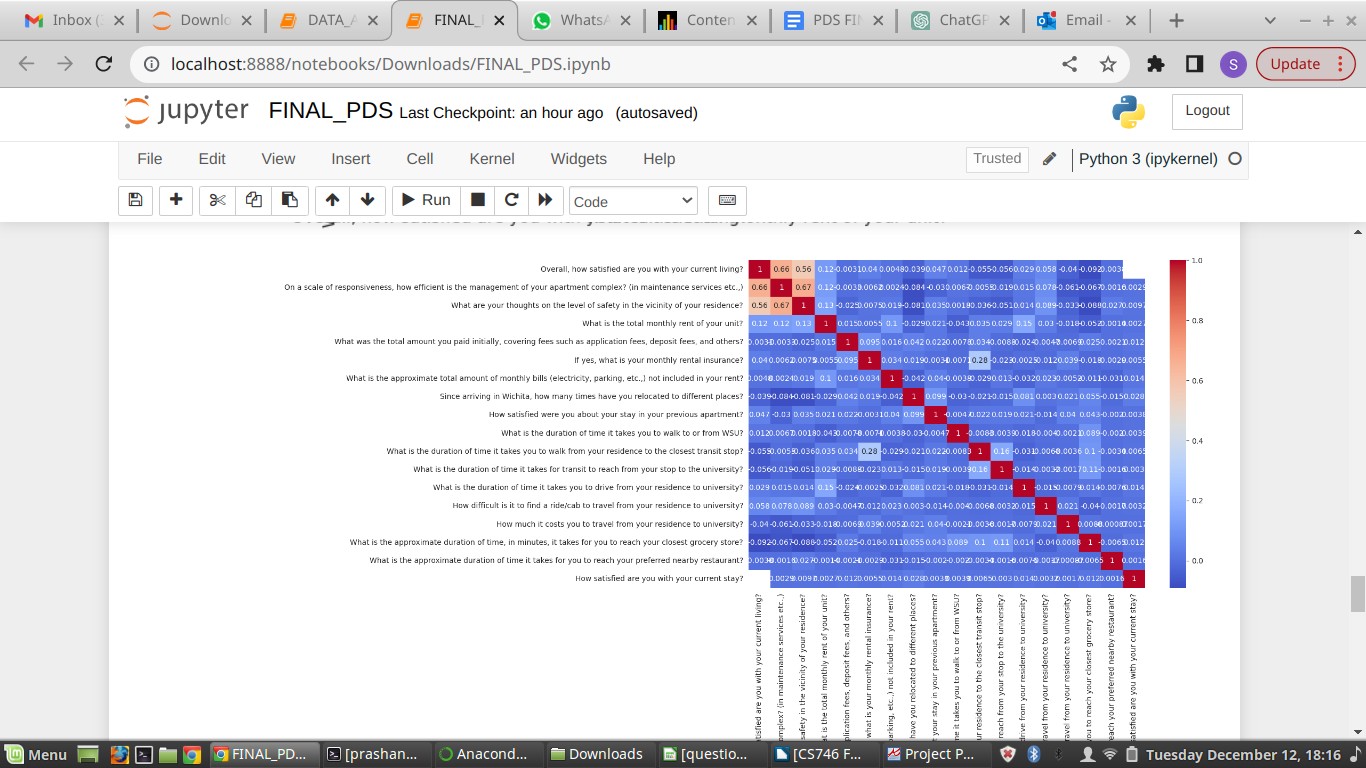
# Project Requirements:

1. We were asked to choose a dataset out of 5 datasets i.e questionnaire responses.xlsx, Apartments spread sheet PDS task.xlsx,transit\_stop.xlsx, restaurants.xlsx, Independent Houses.xlsx. I choose questionnaire responses.xlsx dataset because it has large data when compared to other datasets.Other ones have less rows and columns.
2. **Performing data cleaning or wrangling** on the chosen dataset,making sure of removing all the redundancy in the data,filling all the missing values and finally standardizing the data.We also performed data cleaning on the independent houses.xlsx simultaneously.The size of the dataset got reduced after performing data cleaning.
3. **Performing Exploratory data analysis** on the chosen data. Removing all the outliers if there are any,removing all the null values if any.Seeing the correlation between all the variables in the dataset by correlation matrix, correlation heatmap.Comparing different variables by segregating numerical and categorical ones.

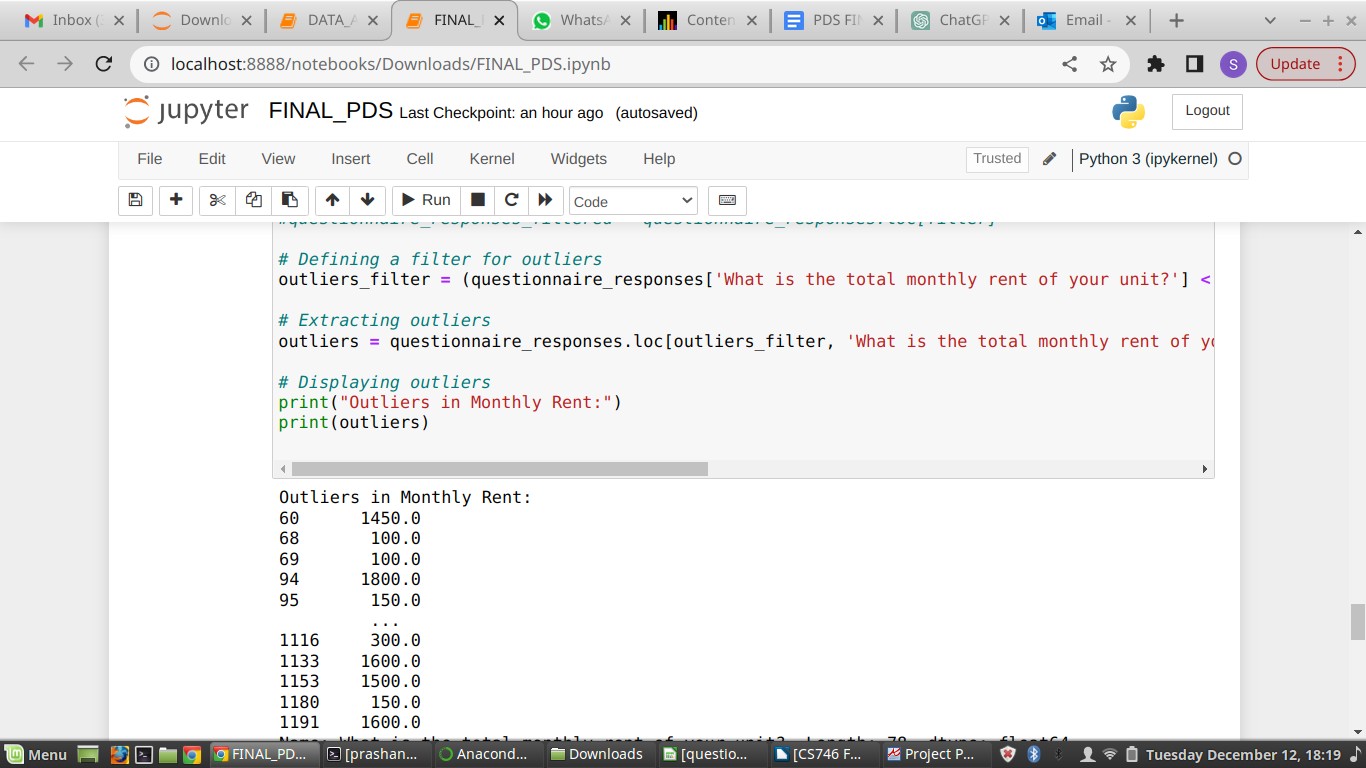
**Correlation\_Matrix:**



**Corelation\_HeatMap:**



## Output for outliers:



1. We performed **Linear regression** on the dataset by choosing different target variable and trying out different predictors. We build different models and print OLS Regression results.
   1. In this project we have choosen ‘What is the total monthly rent of your unit?’ as the **Target variable** and 'Overall, how satisfied are you with your current living?' as the **predictor variable** and performed linear regression . We got output for the model1 as ,

OLS Regression Results

===============================================================================

=====================================

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dep. Variable: | Overall, how satisfied are you with your current living? | R-squared: | | 0.013 |
| Model: | OLS Adj. R-squared: | 0.012 | |  |
| Method: | Least Squares F-statistic: | 15.77 | |  |
| Date: | Tue, 12 Dec 2023 Prob (F-statistic): | | 7.58e-05 | |
| Time: | 17:54:34 Log-Likelihood: | | -1573.3 | |
| No. Observations: | 1199 AIC: | | 3151. | |
| Df Residuals: | 1197 BIC: | | 3161. | |
| Df Model: | 1 | |  | |
| Covariance Type: | nonrobust | |  | |

===============================================================================

=================================

coef std err t P>|t| [0.025 0.975]

const 3.7199 0.109 34.073 0.000 3.506 3.934

What is the total monthly rent of your unit? 0.0004 0.000 3.971 0.000 0.000 0.001

==============================================================================

Omnibus: 154.247 Durbin-Watson: 1.972

Prob(Omnibus): 0.000 Jarque-Bera (JB): 216.626

Skew: -0.966 Prob(JB): 9.12e-48

Kurtosis: 3.778 Cond. No. 4.05e+03

==============================================================================

* 1. We also choose 'What is the approximate total amount of monthly bills (electricity, parking, etc.,) not included in your rent?' as the **Target variable** and 'Overall, how satisfied are you with your current living?' as the **predictor**. We got output for the model2 as ,

OLS Regression Results

===============================================================================

=====================================

Dep. Variable: Overall, how satisfied are you with your current living? R-squared: 0.000

Model: OLS Adj. R-squared: -0.001

Method: Least Squares F-statistic: 0.004094

Date: Tue, 12 Dec 2023 Prob (F-statistic): 0.949

Time: 17:54:34 Log-Likelihood: -1581.1

|  |  |  |
| --- | --- | --- |
| No. Observations: | 1199 AIC: | 3166. |
| Df Residuals: | 1197 BIC: | 3176. |
| Df Model: | 1 |  |
| Covariance Type: | nonrobust |  |

===============================================================================

===============================================================================

====================

coef std err t P>|t|

[0.025 0.975]

const 4.1428 0.039 107.524 0.000

4.067 4.218

What is the approximate total amount of monthly bills (electricity, parking, etc.,) not included in your rent?

-5.492e-06 8.58e-05 -0.064 0.949 -0.000 0.000

==============================================================================

|  |  |  |
| --- | --- | --- |
| Omnibus: | 155.215 Durbin-Watson: | 1.948 |
| Prob(Omnibus): | 0.000 Jarque-Bera (JB): | 217.554 |
| Skew: | -0.977 Prob(JB): | 5.74e-48 |
| Kurtosis: | 3.731 Cond. No. | 661. |

==============================================================================

* 1. We also choose 'On a scale of responsiveness, how efficient is the management of your apartment complex? (in maintenance services etc.,)' as the **Target variable** and 'How many people are staying in your unit?' as the **predictor**. We got output for the model3 as ,

OLS Regression Results

===============================================================================

======================

Dep. Variable: How many people are staying in your unit? R-squared: 0.000

|  |  |  |
| --- | --- | --- |
| Model: | OLS Adj. R-squared: | -0.001 |
| Method: | Least Squares F-statistic: | 0.05309 |
| Date: | Tue, 12 Dec 2023 Prob (F-statistic): | 0.818 |
| Time: | 17:54:35 Log-Likelihood: | -2048.5 |
| No. Observations: | 1193 AIC: | 4101. |
| Df Residuals: | 1191 BIC: | 4111. |
| Df Model: | 1 |  |
| Covariance Type: | nonrobust |  |

===============================================================================

===============================================================================

==============================

coef std err t P>|t|

[0.025 0.975]

const 4.1340 0.169 24.509

0.000 3.803 4.465

On a scale of responsiveness, how efficient is the management of your apartment complex? (in maintenance services etc.,) 0.0093 0.040 0.230 0.818 -0.070 0.089

==============================================================================

Omnibus: 12.144 Durbin-Watson: 1.818

Prob(Omnibus): 0.002 Jarque-Bera (JB): 18.533

Skew: -0.020 Prob(JB): 9.45e-05

Kurtosis: 3.609 Cond. No. 19.0

==============================================================================