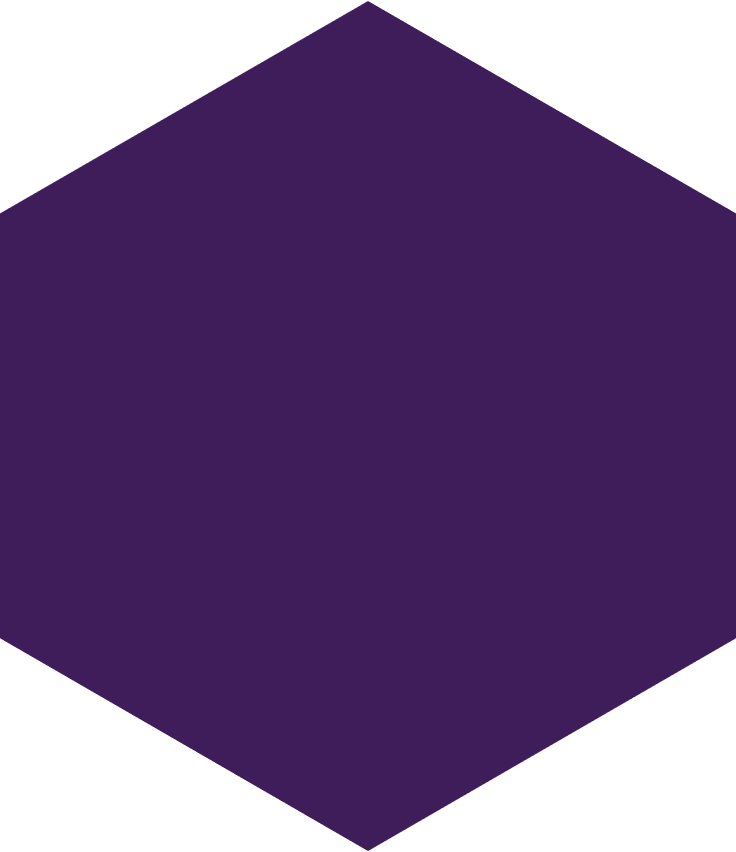
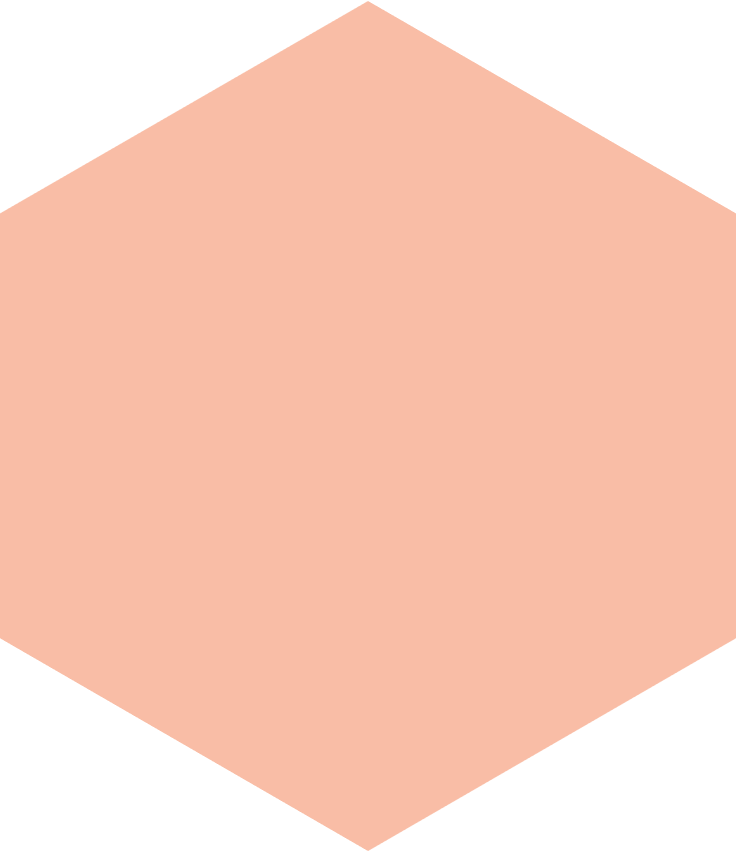


|  |
| --- |
| **[Is this your ideal home?]** |
| [ISSS616 Applied Statistical Analytics with R] |
| This Project Report is partial fulfilment of the academic requirements of ISSS616 Applied Statistical Analytics with R module.  Team Mates: Lee Meng Yong, Ang Hai Soon, Alvina Lee Hui Shan |
|  |





|  |  |  |
| --- | --- | --- |
| [Is this Your Ideal Home?] |  | [What do you consider when you are looking for your dream home? ] • • • |
| [What do you consider when you are looking for your dream home? ] *[Where is a good location?]**[Will my children be able to get into a school easily?]**[How much savings do I need before I can purchase my dream home?]* |
| Section 1: Abstract |
| Singapore is a metropolitan city with limited land space for her citizens. Further with the increasing living costs, there are many considerations that one will have as they look forward to settling down with their life partner and embark on their family planning journey.  In Singapore, Housing Development Board released an annual report, indicating that more than 80% of the citizens are living in HDB flats with a 22.5% increment of new applications in AY 2017 as compared to AF 2016. Therefore, given the large proportion of citizens living in HDB, it will be beneficial if one can refer to one information source before they proceed to look for their dream home.  This project attempts to explore via existing databases to piece together a data product that will be useful for an individual that is either newlywed or about to start a family. Informative and predictive components are the key aims of this data product.  Through linear regression modeling, it was discovered that while there is no significant relationship between overall dataset of schools and resale price, there is significant relationship between certain schools and resale price. Further the top three variables that has a significant relationship with resale price is flat type, flat model and floor square area. |

Section 2: Overall Concept

Given that 82% of Singapore residents stays in HDB[[1]](#footnote-1), housing information is of high interest to an individual residing in Singapore.

Therefore, the purpose of this project is to provide a one-stop information source for all Singaporeans to understand the following:

1. If indeed the price of housing will increase when the location has easy access to certain primary schools
2. Financial considerations that one will need to have before the purchase of a HDB flat.

Given the above, our target audience will be either the newlyweds or those embarking on family planning, i.e. to have a child.

For the newlyweds, financial constraint is a key factor. One needs to assess his or her own earning ability before the housing commitment. Hence it will be in one’s benefits to understand the different housing prices in various planning zones of Singapore.

For those who are married and intends to embark on their family planning journey, given the level of academic competitiveness in Singapore, parents-to-be will assess if indeed there is a need to find out about housing that are near to certain schools, and if they can afford the financial liabilities before they consider a new nest for their family.

A data product created with the above purpose for the target audience will be of positive impact.

Section 3: Data Sources

In order to meet the purpose of this project, we will be using related data from various public sources, including but not limited to:

1. HDB Resale[[2]](#footnote-2) transaction prices from Data.gov.sg
2. List of primary schools in Singapore from Data.gov.sg[[3]](#footnote-3)
3. School details from Ministry of Education [[4]](#footnote-4)
4. Geolocational data from Onemap.sg API[[5]](#footnote-5)

**Housing prices in Singapore -** To understand the prices of resale flats in Singapore, we will be using transaction prices from Data.gov.sg. We will be limiting the data to transactions after January 2016 to ensure the relevancy and recency of our findings.

**Geolocational data -** To obtain geolocational data, we are planning to build a web scraper that connects to Onemap.sg API to extract a list of all latitude and longitude of all resale flats and education facilities in Singapore. With geolocational data, we will be able to triangulate the distance between HDBs and the schools in Singapore.

**Attributes of educational institutions –** While there is no official definition of a good school by Ministry of Education and school ranking data is not public knowledge, we based the selection of school upon the following considerations that parents could have:

1. If it is a primary school with affiliation to another secondary school
2. If it is offering the following programmes:
   1. Government Aided
   2. Government Aided; SAP
   3. Government
3. Whether the primary school is Co-Ed or Single gender

Section 4: Specific Methodology and Creation of Data Product

Given that there are 124033 observations and 36 variables, it is necessary to narrow down to the variables that depicts significant relationship with the key target variable which is resale price. In order to achieve this, linear regression model is implemented for the following variables:

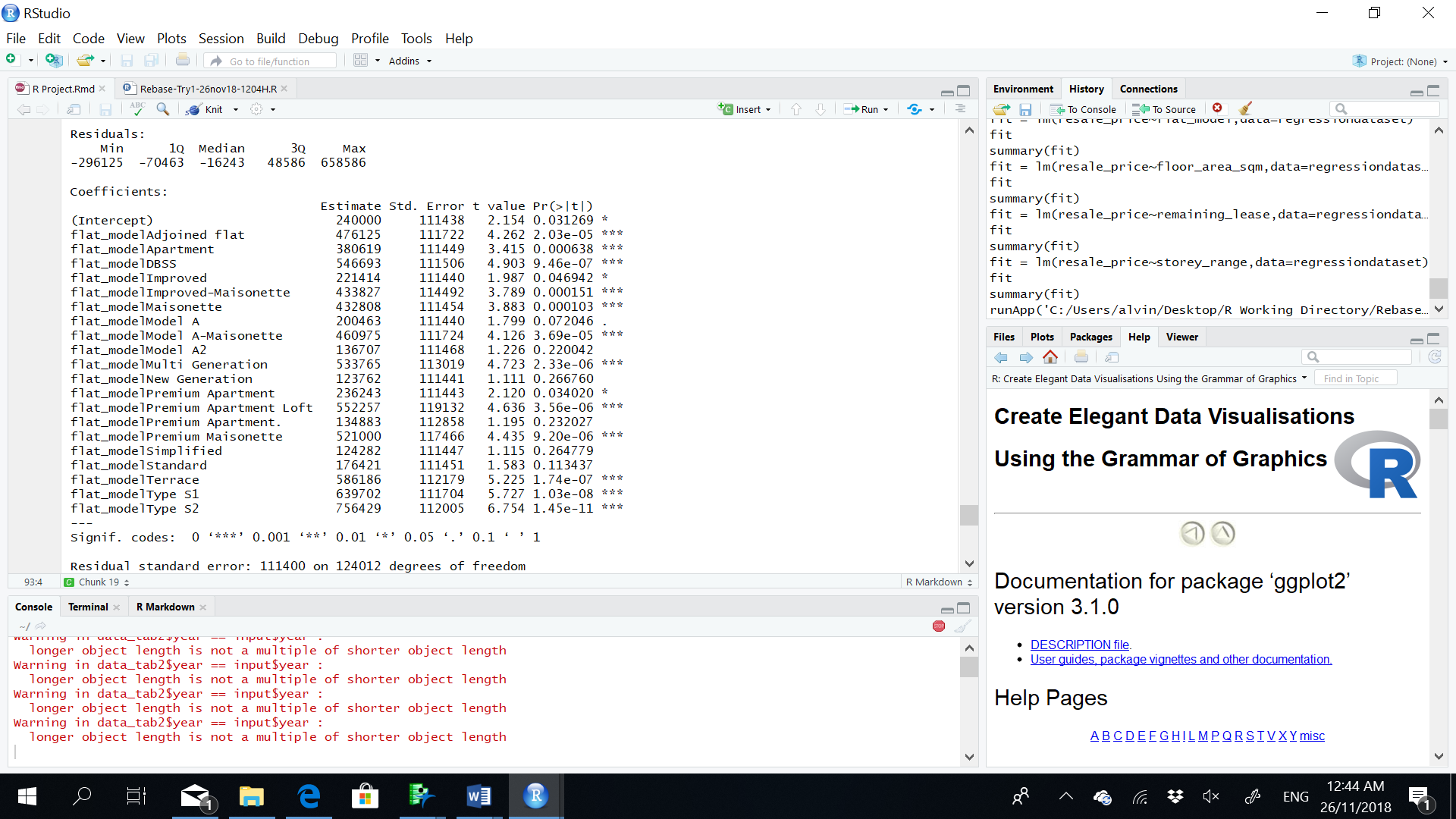
1. Closest\_primary\_school
2. Town
3. Flat\_type
4. Flat\_model
5. floor\_area\_sqm
6. remaining\_lease
7. storey\_range

The following results is obtained:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **R-Squared** | **Residual standard error** | **P-Value** | **Adjusted R-Squared** |
| Closest\_Primary\_school | 0.2439 | 118500 on 123851 degrees of freedom | 2.2e-16 | 0.2428 |
| Town | 0.139 | 126400 on 124007 degrees of freedom | 2.2e-16 | 0.1388 |
| Flat\_Type | 0.4612 | 99970 on 124026 degrees of freedom | 2.2e-16 | 0.4611 |
| Flat\_Model | 0.3306 | 111400 on 124012 degrees of freedom | 2.2e-16 | 0.3305 |
| Floor\_area\_sqm | 0.4548 | 100600 on 124031 degrees of freedom | 2.2e-16 | 0.4548 |
| Remaining\_lease | 0.1353 | 126600 on 124031 degrees of freedom | 2.2e-16 | 0.1353 |
| Storey\_range | 0.1449 | 125900 on 124008 degrees of freedom | 2.2e-16 | 0.1448 |

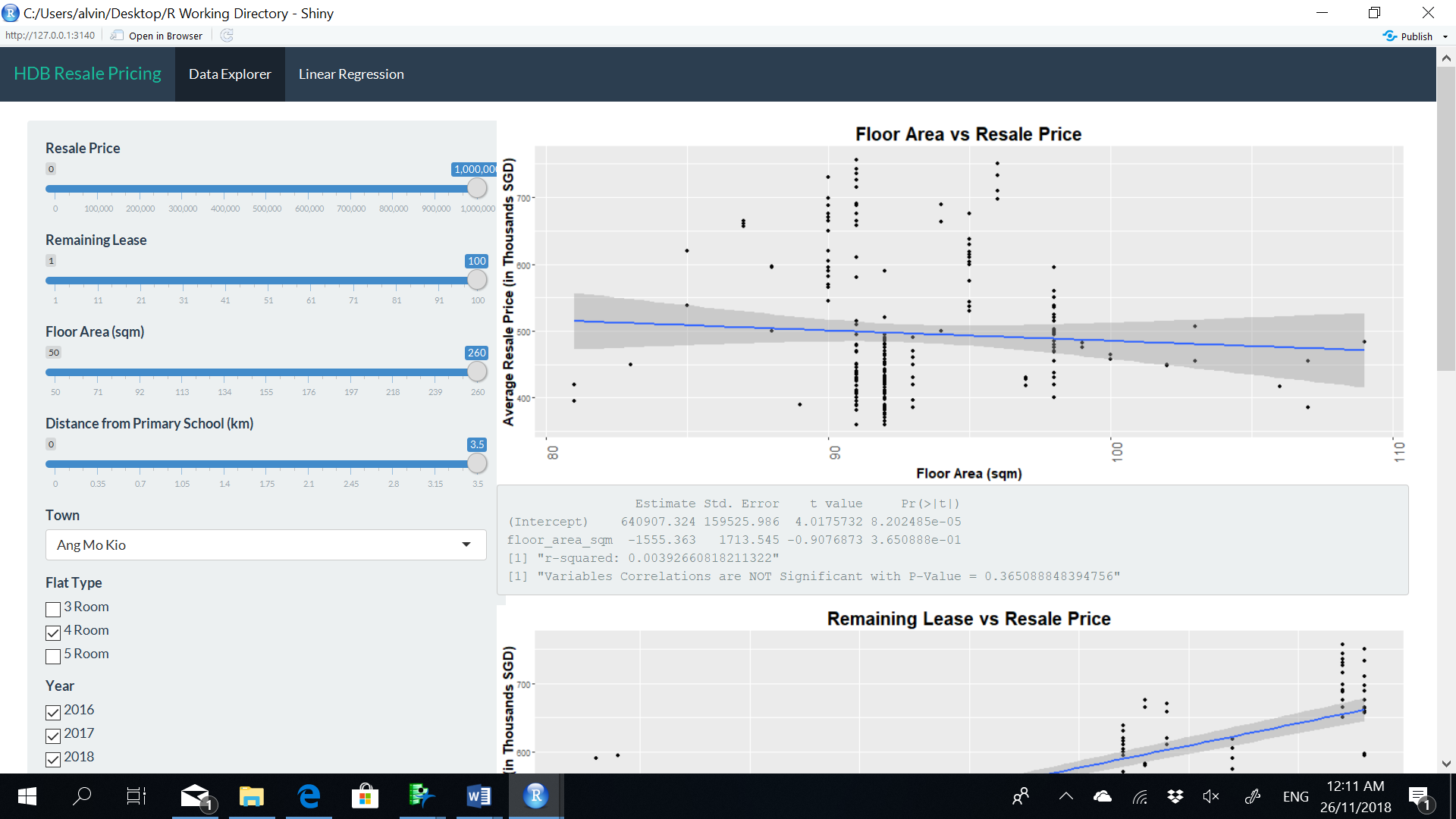
The team plans to build explanatory model using the selected variables in blue as a basis to first explain to the users the pricing differences in various planning regions and the possible correlations with the selected variables such as availability of primary schools, flat type, flat model and floor area. This will provide guidelines on pricing decisions.

It is important to be aware that within each variable, there is further breakdown to depict which sub-category has significant relationship with the target variable.

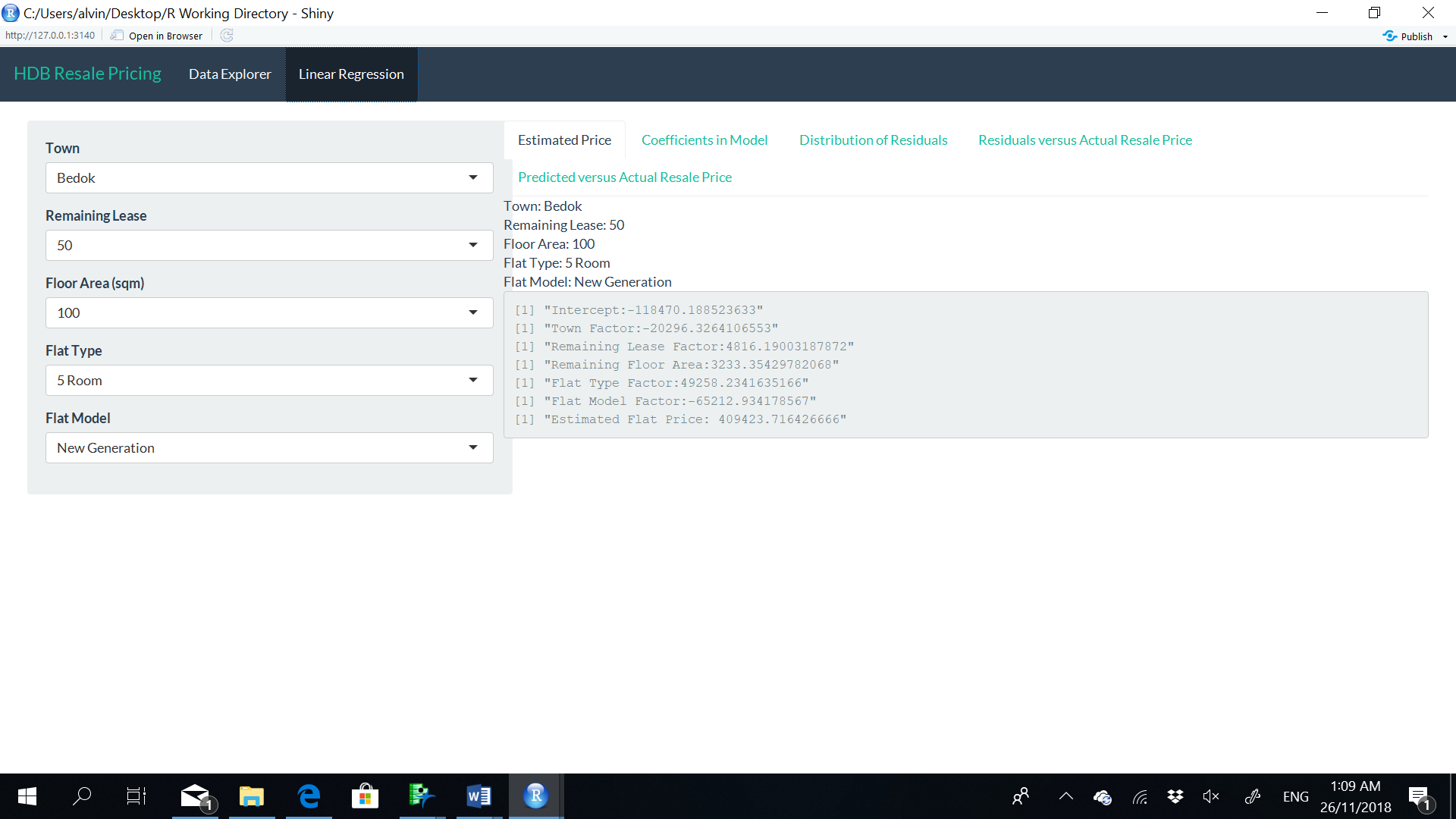


For example, in the above image, not all flat model-sub-category is significant when regressed against the target variable.

Given the above, we also build our explanatory model (also known as the Data Explorer in our data product) to further allow the user to examine the relationship between the average resale prices and the breakdown of variables into sub-category such as 3-room, 4-room etc.



Finally, we also build a predictive model so that the users can base on the selected variables, find out what is the estimated amount of money they will require in order to procure a HDB in the area with the required attributes.



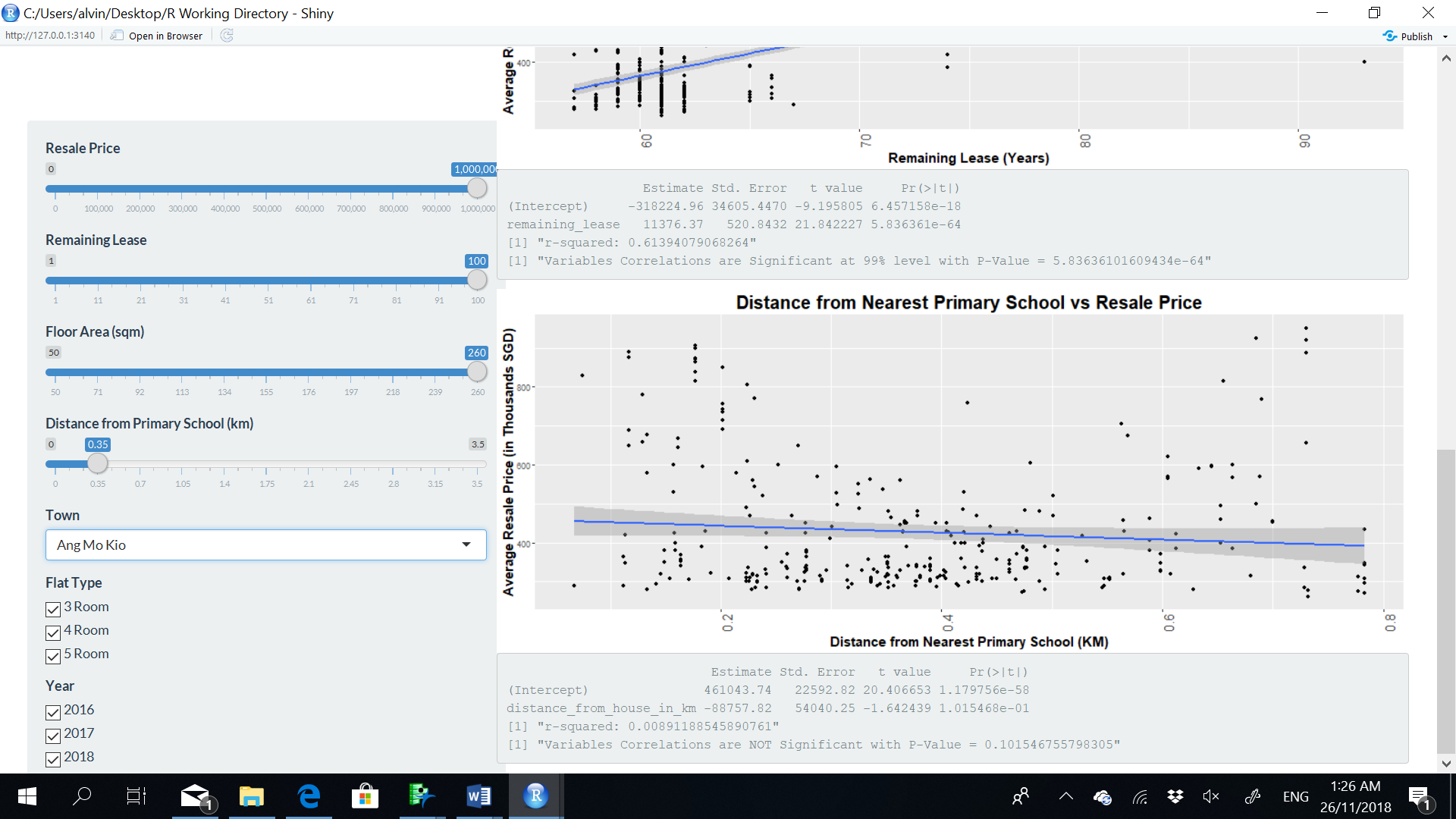
Section 5: Findings

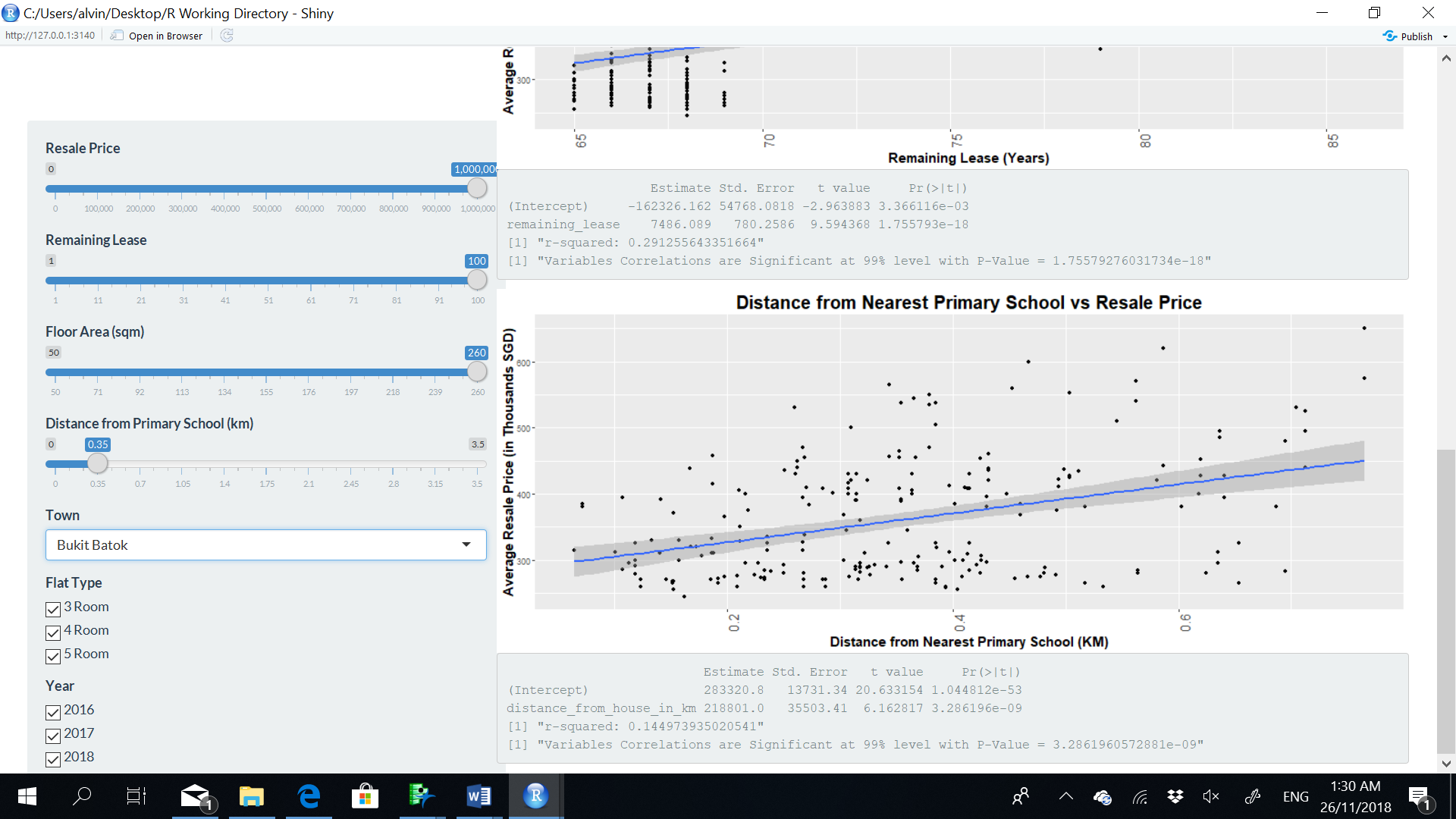
Our initial null hypothesis are as follows:

1. Resale price of HDBs will be affected by the distance between the HDBs and the primary schools
2. Resale price of HDBs will be affected by the choice of flat models
3. Resale price of HDBs will be affected by the floor size of the HDB.

Correlations between Closest Schools and Location of HDB

It is stated in section 4 that not all schools have a significant correlation with the resale price of HDB. With further analysis between the variable “Town”, “Closest\_primary\_schools” and resale price, we observed the following:

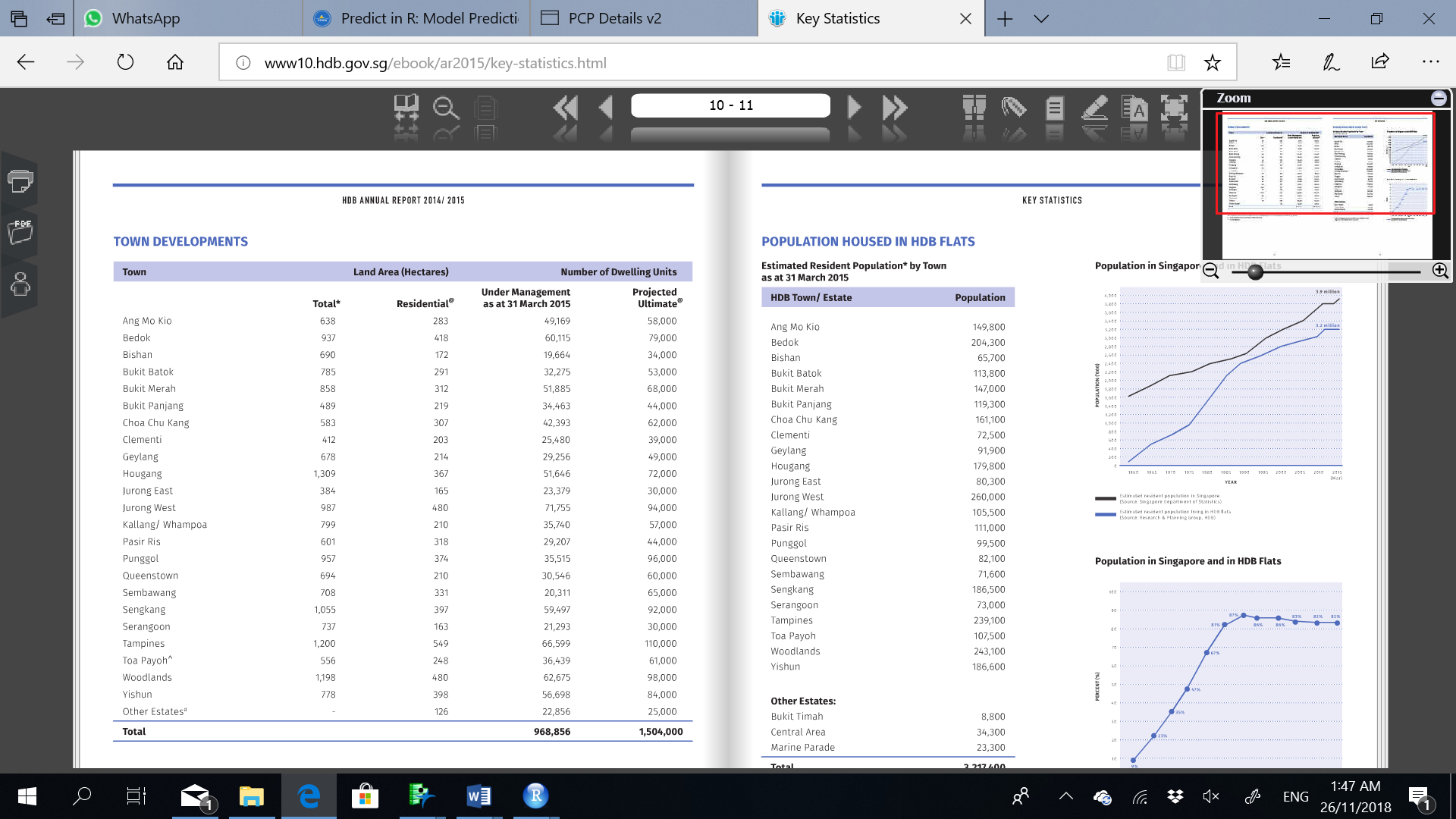




Comparing the two results between Ang Mo Kio and Bukit Batok, the former turned out to be not significant while the latter is. Upon further examination, those towns with a higher number of primary schools (i.e. more than 5) and higher population density tend to depict a significant correlation to the target variable (i.e. resale price). In additional, those towns with popular primary schools such as Tao Nan Primary, Catholic High (Primary) etc., will also depict a significant relationship.

The breakdown of the towns that have similar results as Bukit Batok are as follows:

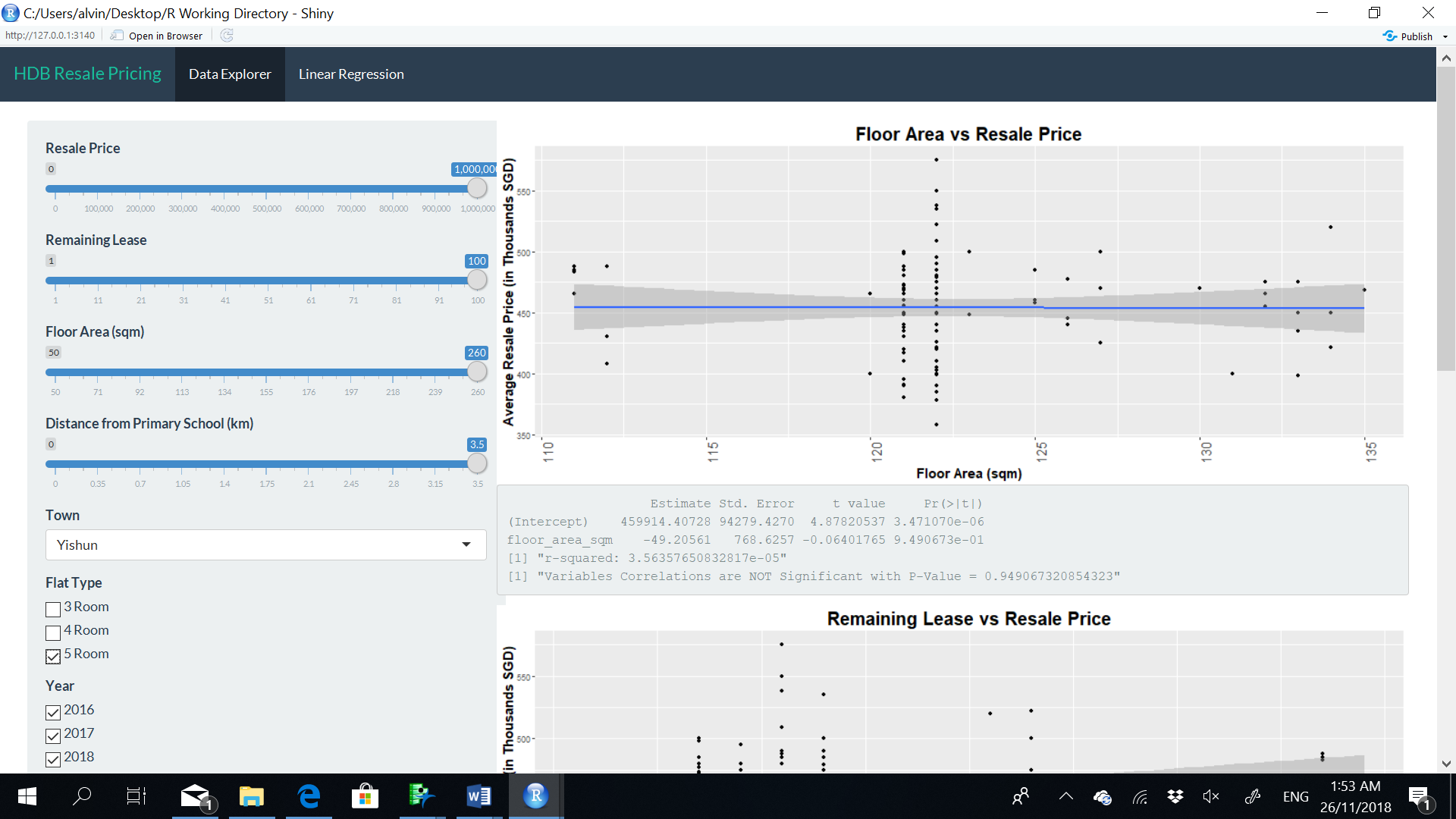
1. Bishan ~ with popular primary schools
2. Bukit Batok
3. Bukit Merah
4. Marine Parade ~ with popular primary schools
5. Pasir Ris
6. Sembawang
7. Tampines
8. Toa Payoh
9. Woodlands
10. Yishun

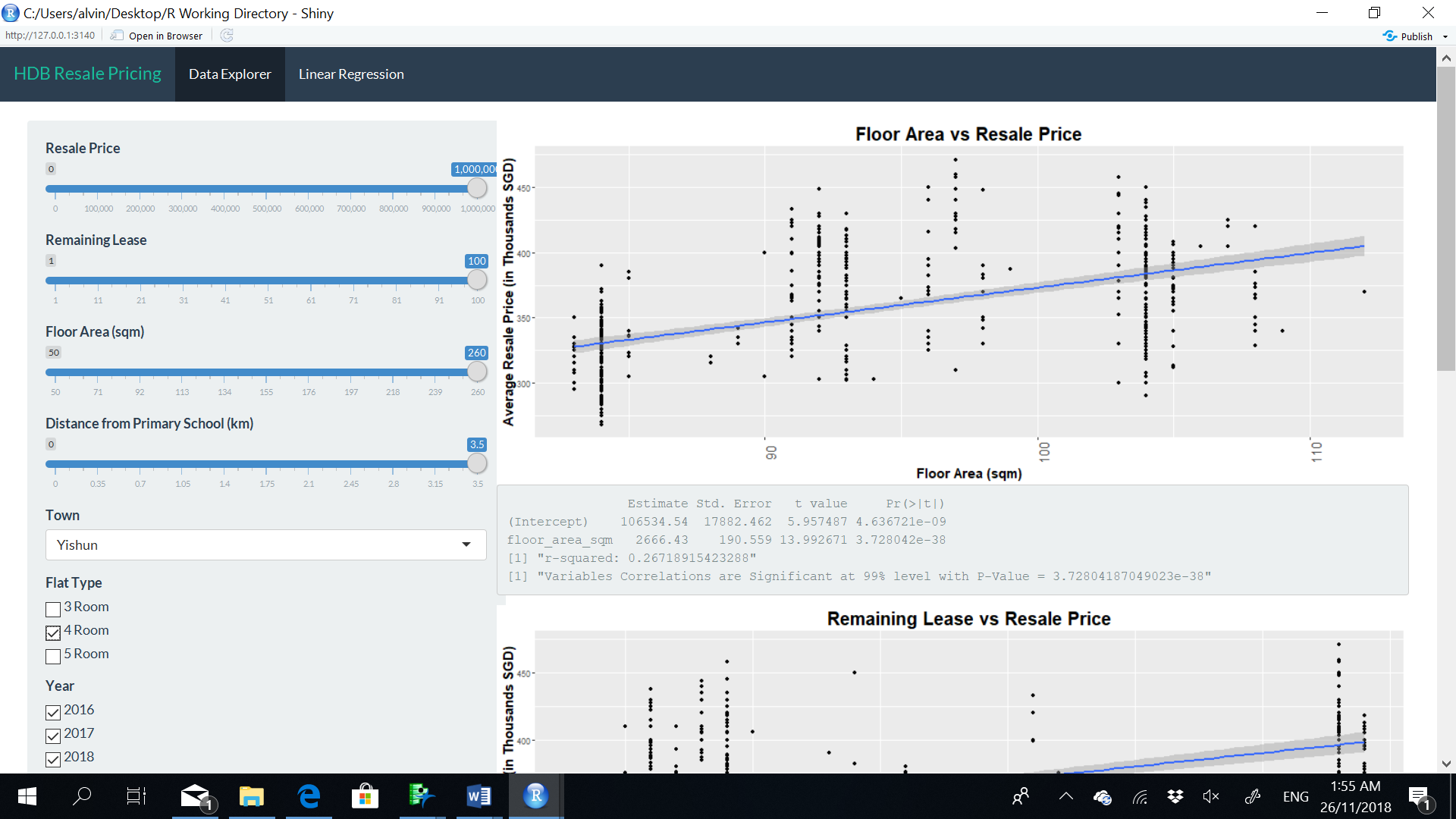


It is interesting to note that while top primary schools such as Nanyang Primary is located in Bukit Timah, the significant correlation is not there. This could be due to the fact that there are not many HDBs in Bukit Timah. It will not be surprising to discover that the pricing of the private housing might be affected by their distance to Nanyang Primary.

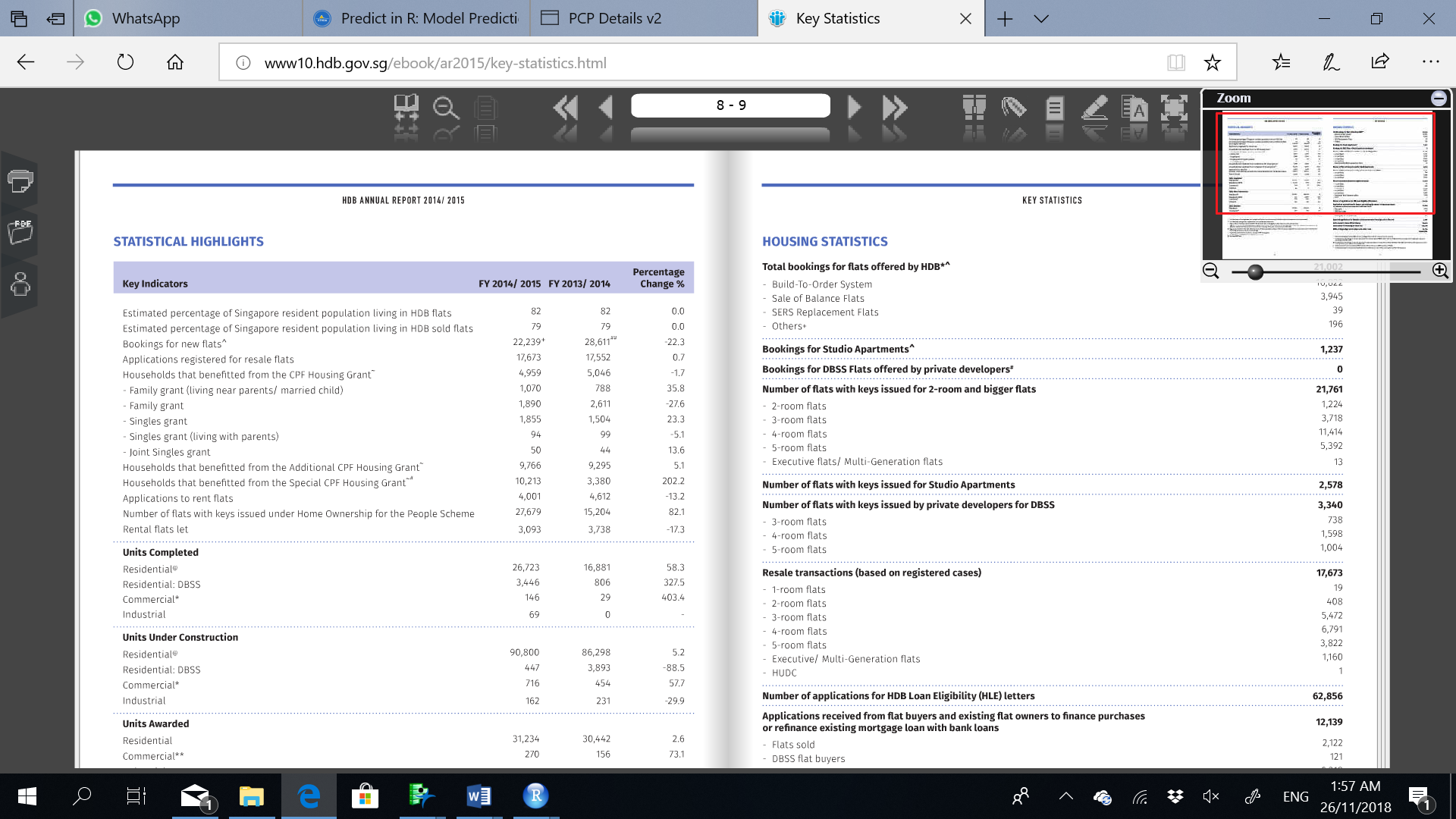
Correlations between Flat Model, Floor area and Location of HDB

It is interesting to observe that generally 5-room flat prices do not have a significant relationship with the Floor Area, while 3/ 4 room flat does.





This is probably because nuclear family model is the general trend in Singapore, hence citizens tend to go for the smaller flat model. This is further confirmed by HDB’s 2015 statistics.



Section 6: Conclusion

It is well known that location plays a big part in the determination of housing prices. From our study, we discovered that house prices go beyond just location. Flat models, distance to certain primary schools and even floor area will all affect the resale HDB pricing. Hence it is very important to have a statistically backed model because this would help to provide more explanatory power to the phenomenon observed.

From there, we hope that the data product will prove to be a useful one-stop information point for individuals who planned to get married or couples who are going to embark on family planning to refer to for reference before they proceed to procure their dream home.

Section 7: References

1. Housing Development Board, *Key Statistics Annual Report 2016/2017*, Retrieved October 16, 2018 from <http://www10.hdb.gov.sg/ebook/ar2015/key-statistics.html>

Section 8: Datasets

1. Data.gov.sg
2. MOE.gov.sg
3. Onemap.sg

1. Refer to HDB Annual Report 2016/2017, retrieved from <http://www10.hdb.gov.sg/ebook/ar2015/key-statistics.html> [↑](#footnote-ref-1)
2. <https://data.gov.sg/dataset/resale-flat-prices>; [↑](#footnote-ref-2)
3. <https://data.gov.sg/dataset/list-of-childcare-centres> [↑](#footnote-ref-3)
4. <https://sis.moe.gov.sg/> [↑](#footnote-ref-4)
5. <https://docs.onemap.sg/> [↑](#footnote-ref-5)