Problem 3: Real Case Study

In this problem, read life data from RealDirect was studied and analyzed. Some of the questions to be answered are as follows.

1. Explore its existing website, thinking about how buyers and sellers would navigate through it, and how the website is structured/ organized. Try to understand the existing business model, and think about how analysis of RealDirect user-behavior data could be used to inform decision-making and product development. Come up with a list of research questions you think could be answered by data:

If I were a data scientist at RealDirect I would ask these questions to improve our dataset:

- Which neighborhood has the maximum number of units and how is that proportional to sales price?
- Which neighborhood gets rented the least and why?
- Are there any particular areas users are interested in (Can be analyzed through browsing activity)?

4. Being the "data scientist" often involves speaking to people who aren't also data scientists, so it would be ideal to have a set of com- munication strategies for getting to the information you need about the data. Can you think of any other people you should talk to?

Some people we can talk to other than data scientists are the Real Estate Sales Agents since they would know things that our system hasn't learnt yet. We could also talk to network engineers who can help us in giving data about where the traffic is coming from. We could also talk to people who are selling houses to get an idea of which locale they are moving to.

5. Most of you are not "domain experts" in real estate or online businesses. Does stepping out of your comfort zone and figuring out how you would go about "collecting data" in a different setting give you insight into how you do it in your own field?

Yes. It was a new and learning experience. It gave me a chance to think outside the box. We can now use these methods on something familiar.

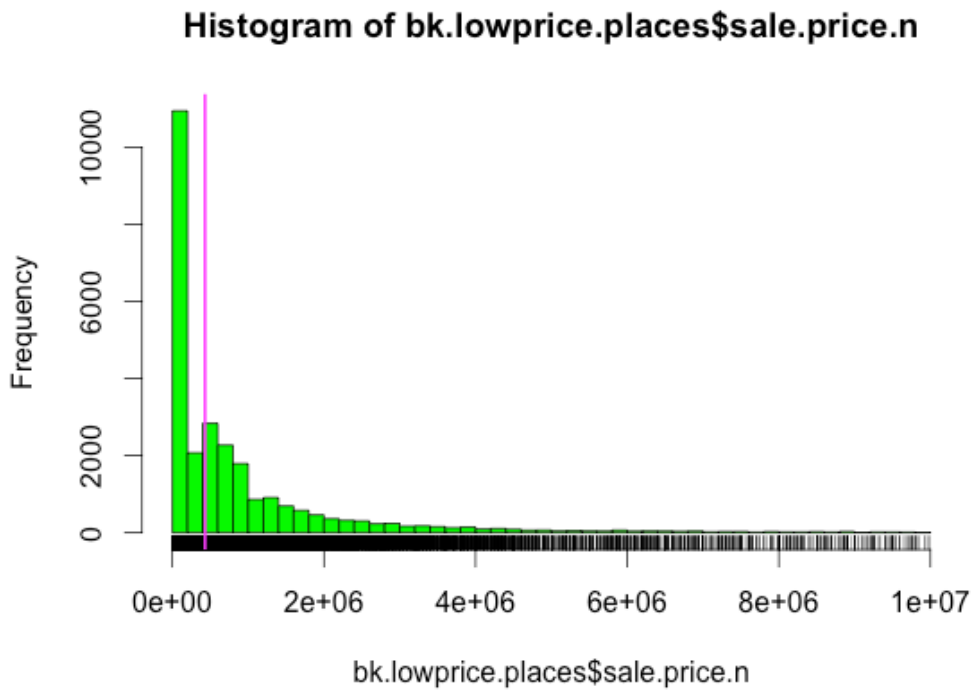
6. Doug mentioned the company didn't necessarily have a data strategy. There is no industry standard for creating one. As you work through this assignment, think about whether there is a set of best practices you would recommend with respect to developing a data strategy for an online business, or in your own domain.

A good data strategy would be to keep all the listings well organized so they are easily accessible by the user. Another point could be user friendly GUI. Too many flashy pop ups or breadcrumbs annoy the general public. Also quick service would be beneficial. For example, if a user is interested in a house, all its details are immediately available and owner is contacted in a timely manner. All the data listings should get updated periodically.

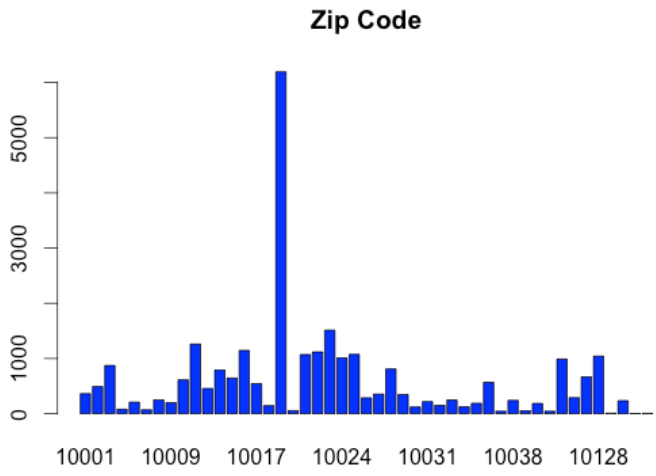
3. Summarize your findings :

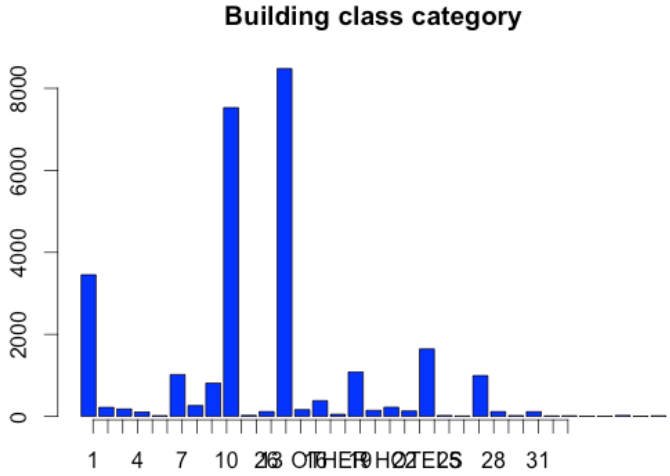
Falguni Bharadwaj

This graph shows a histogram of low price places available with median in pink.



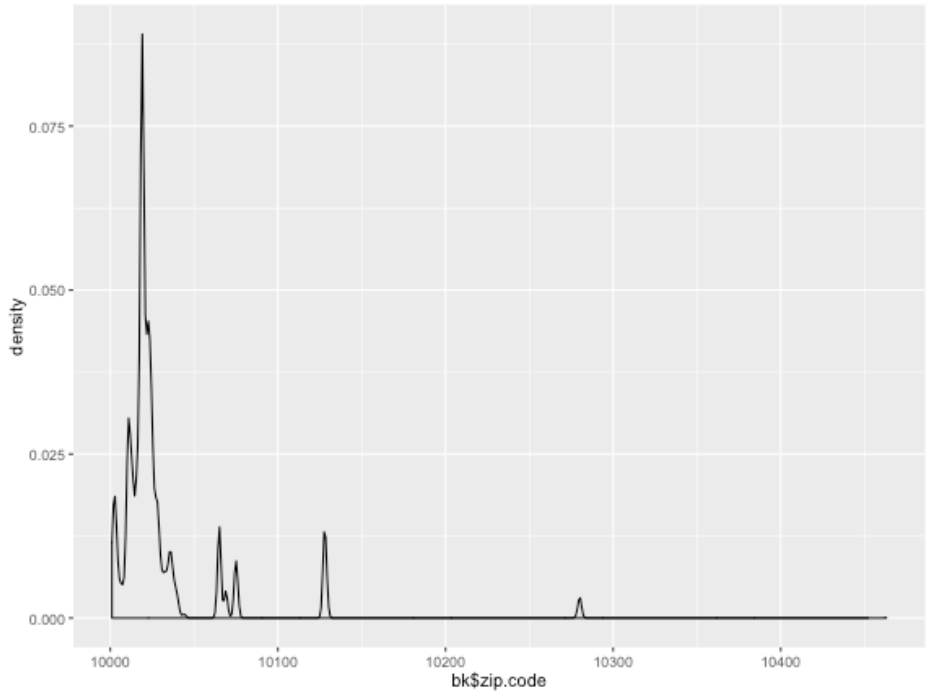
This barplot shows different types of zip codes and the other one shows building class categories.



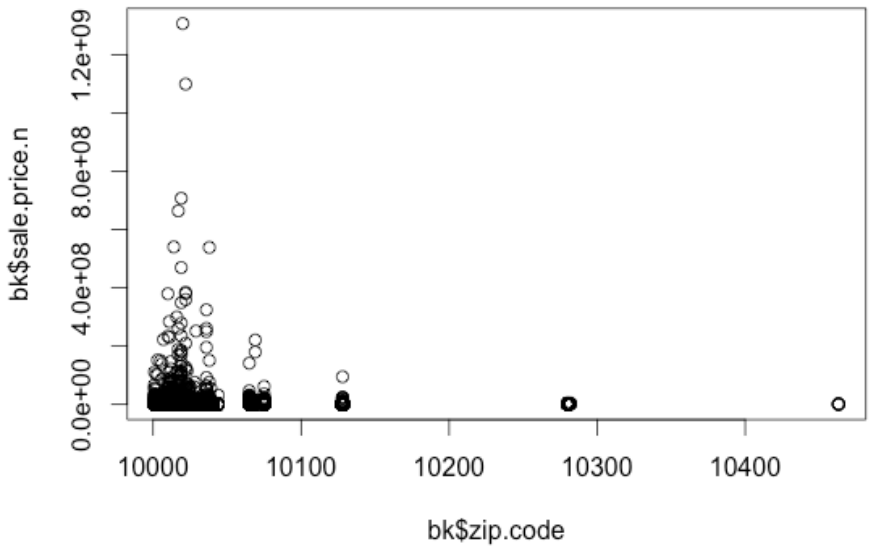


Falguni Bharadwaj

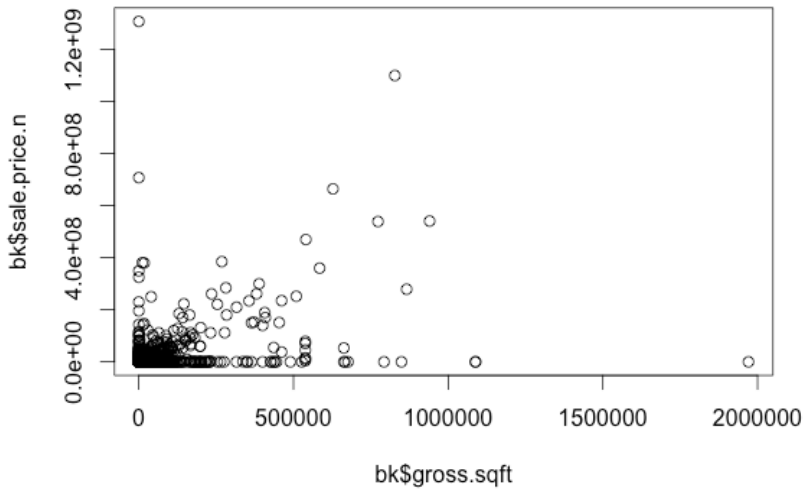
This graph shows distribution of zip code with sale price. It can be inferred that house at zip code between 10000 and 10050 are the most pricey.



Here is another plot of sale price with respect to zip code.

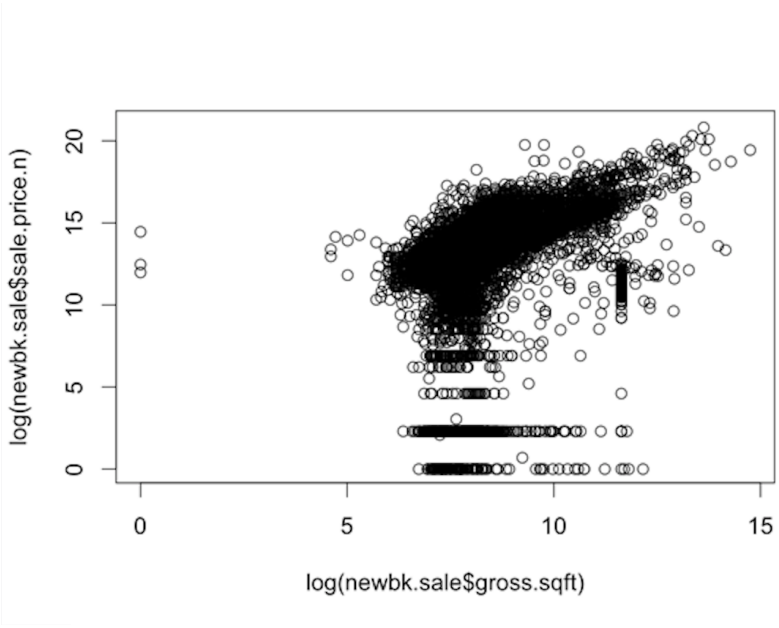


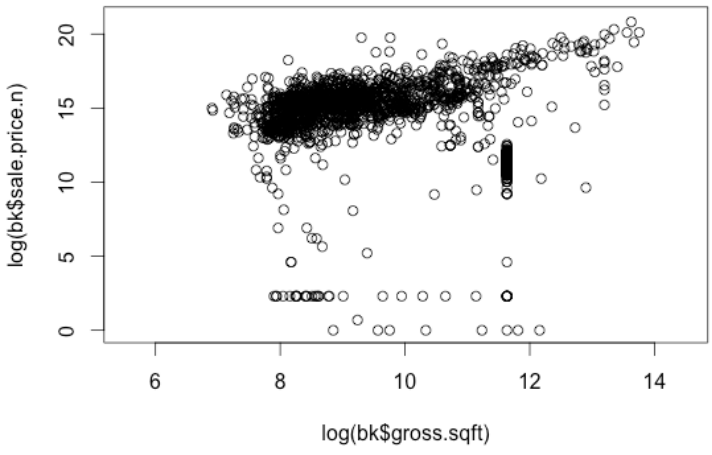
Falguni Bharadwaj



Sale price vs gross sq feet shows outliers in the dataset.

Analysis on multiple data for all boroughs shows for what area how much will the house sell for





Falguni Bharadwaj