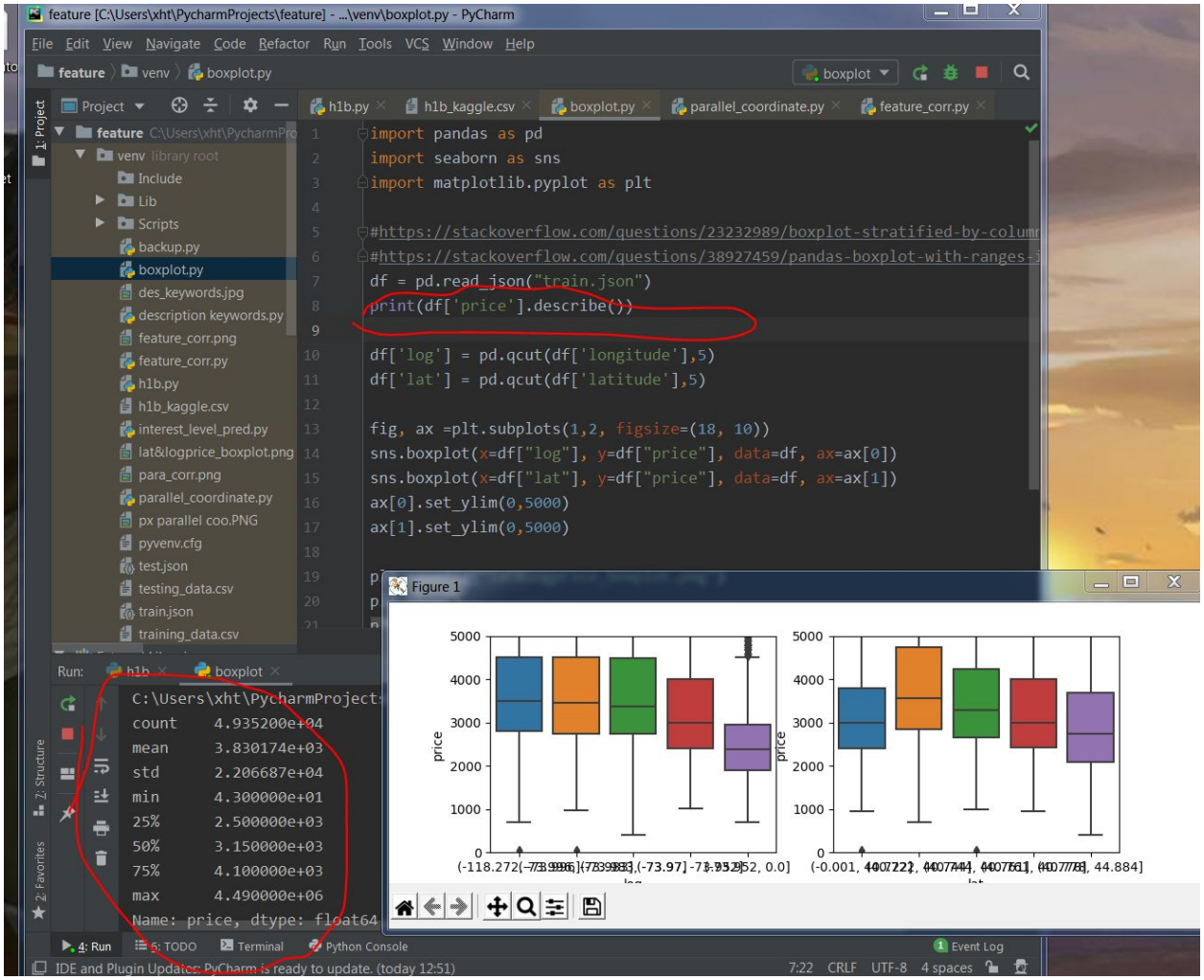
- Developer: Sophia Xiao @ 2021/1/2
- Dataset downloaded from Tianchi

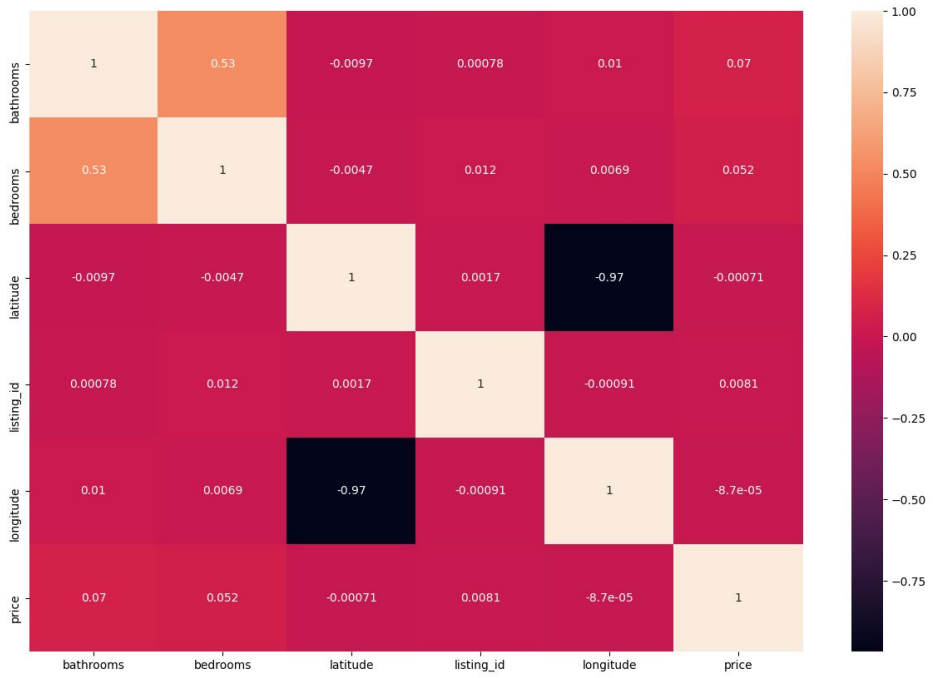
  https://tianchi.aliyun.com/dataset/dataDetail?dataId=83994
- Goal:
  - Boxplot: showing relationship between different latitude/longitude and price
  - Heatmap & parallel coordinate: showing correlation between features
- Website cited
  - Boxplot
    - https://stackoverflow.com/questions/23232989/boxplot-stratified-by-column-in-python-pandas
    - https://stackoverflow.com/questions/38927459/pandas-boxplot-with-ranges-in-x-axis
  - Parallel coordinates plot
    - https://plotly.com/python/parallel-coordinates-plot/
    - https://stackoverflow.com/questions/38103738/plotting-parallel-coordinates-in-pandas-python
  - Heatmap
    - https://medium.com/@szabo.bibor/how-to-create-a-seaborn-correlation-heatmap-in-python-834c0686b88e
- Points:
  - Boxplot:
    - Cut the value of selected col into wanted num of bins
    - Make the figure contains two boxplots (bc lat & logi)

- Set y limit by first look at stat summary of price

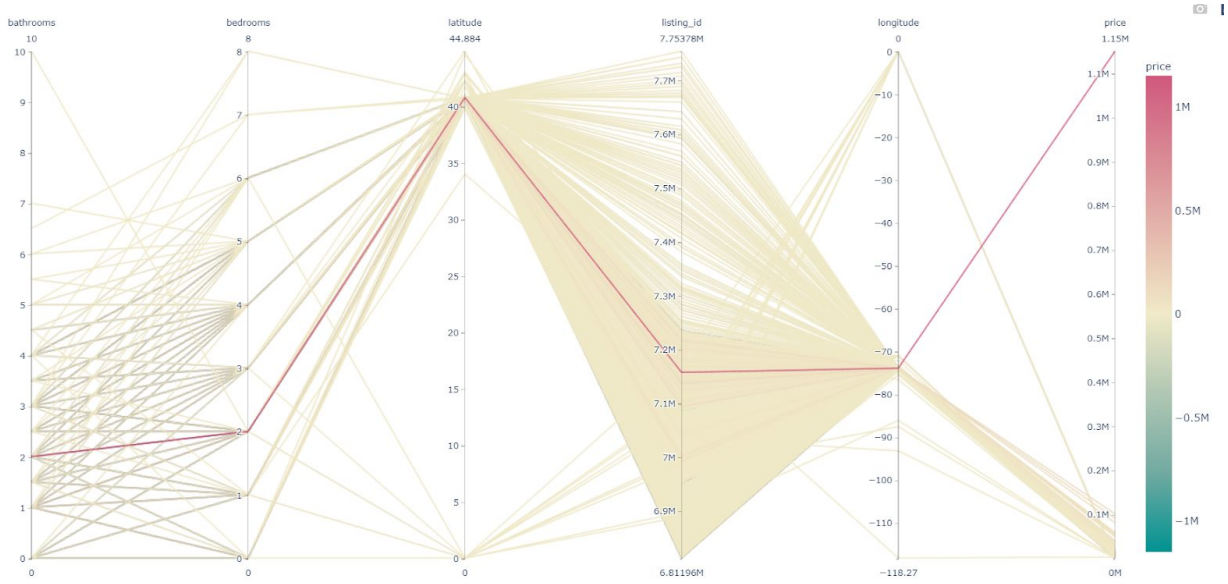


- From the plot we can see the avg rental price is mostly bt 3k - 4k, the higher longitude, the lower price.

- Parallel coordinate plot and heatmap- assign in the right data to graph



- Showing there is not much correlation between features except for longitude and latitude which have a strong negative linear relationship



    - Like indicated in the heatmap, there is not an apparent relationship between each other.

- Problems I had during the project:
    - The different tools (px vs pandas plotting parallel coordinates) produce different plot, but the variables I assign are the same
    - I'm not sure if I really understand how to interpret a parallel coordinate plot