

CISC/CMPE Advanced Data Analytics W22

Exercise 1

January 13, 2022

1 Introduction

The goal of this assignment is to get familiar with DataSpell or other IDE for running Python and R for data exploration. The provided data is a one-month listing data from airbnb, for the city of Montreal. You are welcome to explore the data and play with it, i.e., using it to check and improve your coding skills.

Since this is a rental price data, we also provide the following references (more will be shared in the QAforum if needed):

- <https://towardsdatascience.com/predicting-housing-prices-with-r-c9ec0821328d>
- <https://rpubs.com/jhofman/nycmaps>
- <https://rpubs.com/ablythe/520912>

Submission: you need to submit your analysis as an executable R Markdown file or Python Jupyter Notebook file.

2 Data Exploration and Hypothesis Test

In this exercise, given the provided data, please write code to answer the following questions:

RQ1 (30 points) Provide at least three types of statistical summaries of the listing price (“price” column) in the data set. You can use plot or numeric values. Write a summary in the markdown cell, stating your finding(s) from your provided statistics summaries.

RQ2 (20 points) Design and perform a statistical test involving list room type and list price. In a markdown cell, describe the null hypothesis statement of your test and why you choose a specific statistical test method. You can consider more attributes if you want.

RQ3 (20 points) Design and perform a statistical test involving neighborhood and list price. In a markdown cell, describe the null hypothesis statement of your test and why you choose a specific statistical test method. You can consider more attributes if you want.

RQ4 (30 points) Design and perform at least two correlation analyses among attributes provided in the dataset. Describe why you believe it would be interesting to analyze the correlation between selected attributes. Write code and describe your main findings from the correlation analysis.