

# AirBnB & Zillow Data Challenge

## Challenge and Expectations

Isn't data just more fun when you can interact and play with it? Well, that's exactly how we challenge our data analysts at Capital One. We need to find great people to join our team as we develop software data products across our three key areas of data work:

- » **Builder Mindset:** Leverages creative and adaptive problem solving to selecting the right tool for the job; seeks automated and efficient solutions to manual or repetitive processes.
- » **Data Management:** Strategically leads efforts to systematically evaluate, and document; monitors our data in a sustained and organizationally recognized way »
- Business Intent:** Translates business needs into actionable solutions or data products; effectively communicate results to stakeholders and technical partners.

This challenge is your next step in showing Capital One what you can do. After receiving data instructions you're putting hands-to-keyboard and have 1 week to submit a working data product, per the submission instructions, including:

- ✓ Working code with documentation
- ✓ Documentation of metadata and data quality
- ✓ Visualizations of key insights

Ready to show off your data chops? Let's go!

## Problem Statement and Instructions

### **Problem Statement**

You are consulting for a real estate company that has a niche in purchasing properties to rent out short-term as part of their business model specifically within New York City. The real estate company has already concluded that two bedroom properties are the most profitable; however, they do not know which zip codes are the best to invest in.

The real estate company has engaged your firm to build out a data product and provide your conclusions to help them understand which zip codes would generate the most profit on short term rentals within New York City.

. You will be looking at publicly available data from Zillow and AirBnB:

- Cost data: Zillow provides us an estimate of value for two-bedroom properties •
- Revenue data: AirBnB is the medium through which the investor plans to lease out their investment property. Fortunately for you, we are able to see how much properties in certain neighborhoods rent out for in New York City

Capital One Confidential

- You can assume an occupancy rate of 75% or you can come up with your own model to calculate occupancy; just let us know how you came to that calculation

After meeting with the strategy team, you've got an idea of where to start, key concerns, and how you can help this real estate company with the market data while keeping the following assumptions in mind:

- The investor will pay for the property in cash (i.e. no mortgage/interest rate will need to be accounted for).
- The time value of money discount rate is 0% (i.e. \$1 today is worth the same 100 years from now).
- All properties and all square feet within each locale can be assumed to be homogeneous (i.e. a 1000 square foot property in a locale such as Bronx or Manhattan generates twice the revenue and costs twice as much as any other 500 square foot property within that same locale.)

Capital One Confidential

### **Instructions**

As you start the challenge, realize that this is real-world, imperfect data. We recommend planning about 4 hours to complete the Data Challenge, but it's not timed, and you are judged on the quality of the work submitted. If you find yourself uncertain of what the "right" answer is, use your best judgment, make an assumption (document the assumption), and keep going.

Overall, we first ask you to show your data skills in three areas at a basic level, and then, in the last step, tell us what you would do next to provide a better conclusion.

1. **Quality Check** – bad data is worse than no data at all
  - a. Understand the data while keeping your final output in mind
  - b. Highlight two to three data quality insights based on your analysis of the data
  - c. Create metadata for any derived fields or metrics used to complete your

analysis

**2. Data munging** – get the data

a. The datasets do have different units of time – in order to complete the analysis, you will need to determine a common unit of time

b. Write a function that can link the data together in a scalable way when new data is available or for when we are ready to approach a new market 3. **Craft a visual data narrative** – *Charts and plots must be generated from your code; not from produced in external standalone software like Excel*

a. Visualize metrics for profitability on short term rentals by zip code b. Summarize your key insights and conclusions based on the data and your analysis

4. **What's Next** – We recognize that 4 hours isn't a lot of time... and you've probably come up with a number of great ideas from an analytical or visualization perspective that you don't have time to do. Tell us (but don't do any work) what you would/could do next to inform a better decision or deliver a better product to the real estate company.

## **Data and Tools**

*Solutions that require purchase of a software license or purchased access to data will not be accepted regardless of whether or not Capital One uses said software or data. Abide by all applicable laws and regulations regarding the use of software or external data sources. If you have questions about a particular software package, please contact your recruiter immediately.*

### **Data**

You need Airbnb and Zillow datasets for this Challenge and the datasets can be found here: <https://github.com/CapitalOneRecruiting/DA>. Please note that there are 4 Airbnb datasets that will need to be combined and that the first file contains a header row.

In addition, you should have received the following additional files from Recruiting

Resource	You should see
----------	----------------

Capital One Confidential

This document	AirBnB_Zillow – Data Challenge.docx
Technical Considerations	Data_Challenge_Technical_Considerations.html
Metadata	AirBnB_Zillow – Metadata.docx

### **Tools**

Here are some example platforms you should feel free to use. By no means are you limited to this list, and our solution review team will be able to evaluate solutions in most languages. If you really do have a question about the platform you would like to use to solve the problem, contact your recruiter with the exact setup you'd like to use (including OS and specific versions when

applicable), your backup choice, and they can seek verification for the platform

Platform example Notable packages
Anaconda Python Distribution notebook, pandas, matplotlib, bokeh
R R, Shiny, plyr, ggplot
Javascript D3, nvd3, node.js, Tableau
Java virtual machine Groovy, Scala
Other software packages with which you are familiar

## **How to submit**

**Congratulations on completing the Data Challenge! Please see the following instructions for how to submit your work.**

Submission is easy – just email to [dataanalysisrecruitingmailbox@capitalone.com](mailto:dataanalysisrecruitingmailbox@capitalone.com) a single ZIP file (< 10 MB) containing:

1. Working source code file with documentation
  - ☐ Code
  - ☐ Source documentation (e.g., a README file)
  - ☐ Any generated graphics files
  - ☐ *If you added data:* if you added more than a couple of MB of data, provide a program or script, with documentation, to download the data set
2. Documentation including metadata for any data created and your data quality insights
3. Visualizations and key insights from those visualizations

**Please do not post your code or documents to any public repositories.** Capital One

## **Acknowledgements**

The data for this challenge were sourced from:

- [Zillow Group, Inc. \(2016\)](#)
- [Airbnb](#)

To maintain the integrity of the assessment process, please complete the assessment independently, keeping all content confidential. Feel free to consult resources, including people and documents on and off the internet, for advice and examples. However, all code delivered (excluding your packaged dependencies) must be solely your independent work. Using others' work with or without permission, or failure to maintain the confidentiality of the assessment process and materials, will disqualify you from further consideration and other consequences, up to and including termination of employment.

The assessment is personalized for your use only and should not be accessed by, or redistributed to anyone, as all content including instructional material is proprietary and strictly confidential. Please reach out to our Assessment Support Team at [careers@capitalone.com](mailto:careers@capitalone.com), if you're experiencing technical issues or have any questions.

Best of luck,  
Capital One Recruiting



If you require a reasonable accommodation to take the assessment, please contact Capital One Recruiting at 1-800-304-9102 or [RecruitingAccommodation@capitalone.com](mailto:RecruitingAccommodation@capitalone.com). All information provided will be kept confidential and will be used only to the extent required to provide a reasonable accommodation.

Confidential