**You are free to structure your presentations to your liking, but students tend to have success with the following format.**

**Title Slide**

**Include the name of the Project and Group Members**

Leo’s House Hunt….

Freguenet

Dan Hong

Derek Denning

Mike George

**Motivation & Summary Slide**

Leo is becoming increasingly frustrated by our basic github questions and ready to retire from coding. He’s intrigued by Washington DC real estate and looking for where best to buy?

Objective: Where and what to buy for an optimal return in 10 years assuming current market conditions remain stable

**Define the core message or hypothesis of your project.**

**Describe the questions you asked, and why you asked them**

Is there a clear quadrant, or ward that is poised to deliver superior investment returns over the next 10 years?

**Describe whether you were able to answer these questions to your satisfaction, and briefly summarize your findings**

TBD

**Questions & Data**

We used the D.C. Residential Properties dataset from Kaggle as our sole dataset.

**Elaborate on the questions you asked, describing what kinds of data you needed to answer them, and where you found it**

We needed key data parameters that would provide insight into a quadrants or wards investment prospects. Some of the data was numerical (sale px, transaction count) while others were qualitative (city quad / ward). Some of the key function we used to cut our data included groupby, .mean, .count, .agg, .loc. Important parameters in our data include:

* Price Year built Building type
* Sale Date Bedrooms Square feet
* Location (quad/ward)

**Data Cleanup & Exploration**

**Describe the exploration and cleanup process**

We spent most of last Saturdays class just exploring to understand what parameters were available to analyze.

**Discuss insights you had while exploring the data that you didn't anticipate**

1. Upwards of half of the 158,000 records didn’t include price data
2. Commercial properties included in the “Residential” dataset severely skewed our findings and needed to be removed.
3. Data prior to 1992 was too sparse to be meaningful and needed to be removed

When all was said and done, we were left with third of the original data ~58,000 rows

**Discuss any problems that arose after exploring the data, and how you resolved them**

1. Sale date was embedded in a string and we needed extract month and to put together our times series data

housing\_df2['YEAR'] = pd.DatetimeIndex(housing\_df2['SALEDATE']).year

housing\_df2['MONTH'] = pd.DatetimeIndex(housing\_df2['SALEDATE']).month

**Present and discuss interesting figures developed during exploration, ideally with the help of Jupyter Notebook**

**Data Analysis**

**Discuss the steps you took to analyze the data and answer each question you asked in your proposal**

**Present and discuss interesting figures developed during analysis, ideally with the help of Jupyter Notebook**

**Discussion**

**Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?**

**Post Mortem**

**Discuss any difficulties that arose, and how you dealt with them**

We initially discussed what we wanted to answer before exploring the data. Once we explored the data and learned the dataset wasn’t this vast oasis of interesting findings …we needed to regroup as a team and come up with plan B. Worse, as a group, our data exploration evolved into us working on different aspects of the data. We resolved this by having frequent touch points to report finings and ensure we were all marching towards the goal of having something presentable.

**Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?**

It would have been interesting to look at additional parameters such as number of bedrooms/ bathrooms, square footage etc. and extrapolate historical returns forward with regressions. With more time we could have delved deeper into specific neighborhoods instead of staying at the Ward level. It would have also been interesting to overlay additional datasets such as crime and walkability scores to ensure Leo remains safe and gets his steps in.

**Questions**

**Open-floor Q&A with the audience**