**Slide 1 (Jack)**

Overall Objective:

* Identify investment targets in the Houston area based on zip code. Key factors taken into consideration include: number of permits in the area, home sale price, number of active listings, pending ratio, new listings, mortgage tax data, income tax data, real estate tax data, charity tax data, etc.
* What areas have highest increase of price/sqft
* What factors correlate to highest price/sqft increase
* Based on these correlating factors, which areas are good to invest in?

KPIs:

* Cumulative Returns based on listing price per squarefoot
* Correlation of different data points
* Predicted ROI

Assumptions

* We would find strong correlations with the given datasets that we ingested
* Houses sold for listing price (best data source we could find)

**Slide 2 (Vu)**

Data Utilized

* Realtor.com Listing Data
* Permit Data from city of Houston
* Tax data from IRS
* Zip code radius API

Data Not utilized

* Zillow data for listing price
* Crime data from the city of Houston
* Census data from census.gov for permit data
* Schooling data from meech.com, TEA, children at risk

**Slide 3 (Mark)**

Data Cleansing

* Cleaning all data into usable pivot tables based on zip code
* Getting data into format useable for analysis and visualizations (re-configuring columns, re-setting indexes, transposing data sets)

Blockers:

* Finding sufficient data to perform a detailed analysis (schooling data, crime data, census data).
* Finding data in same sample frequency and time range.
* Difficulty finding data with unique identifier (we decided on zip code)

**Slide 4 (Dennis)**

Analysis

* Created a dictionary of dataframes to group data by Zip Code
* Identified the strongest correlation to percent change by Zip Code
* Evaluated trends by linear regression
* Used secondary linear regression to predict future ROIs

**Slide 5 (Dan)**

Visualization

* Map plot of zip codes with cumulative returns
* Overall Active listings per month of entire area
* Average listing price per month of entire area
* Median price per sqft per zip over time
* Top tens
  + Cumulative pct change in average listing price
  + Cumulative pct change in Average income tax
  + Construction Permits
  + Cumulative pct increase in price per sqft
* Heatmap for Correlation
* Bar Chart for predictive ROI
* Scatter plot with linear regression analysis

**After Jupyter Lab Presentation**

Taking it past MVP

* Index by something more friendly (not zip codes)(possibly neighborhood names)
* True MLS data with final sale price instead of listing price
* Multiple linear regressions in the analysis for the ROI predictor
* Additional alternative data sources to take it past the standard housing data (dog valley)
* Give the users their own choose to input radius, include other states/countries