**Capstone Project**

**Project proposal**

**1.** **Group description**

**1.1.** Group name

|  |
| --- |
| Breakout Room 2 |

**1.2.** Students names, background and target industry if any

|  |
| --- |
| Michael McGuigan   * Background in project management and mechanical engineering. Targeting large scale consulting industry in New York City.   Ben Burkey   * Background in market analysis, project management, teaching   Jonah Gerstel   * Background in process engineering, cost and optimization analysis, and project management.   Lena Chretien   * Background in academics (teaching and research), oceanography and meteorology. |

**1.3.** Group structure: roles and responsibilities

|  |
| --- |
| Group Leader and coordination   * Michael McGuigan   Google Cloud expert   * Ben Burkey   Data Wrangler   * Jonah Gerstel   Data cleaning (Python)   * Lena Chretien   Base model and clustering, and modeling by cluster   * Group effort |

**2. Why** do we want to develop a data science project?

**2.1 Objective**: what problem do you want to solve? What questions are you trying to answer? How will you **measure the success** of your analysis from a business/user perspective?

|  |
| --- |
| Our objective is to predict the Zillow Rent Index (ZRI) for multi-family homes in various cities. We plan to start by running a cluster analysis to determine similar markets. This will allow a strong blend between detail and accuracy by having different predictive models based on the cluster analysis results. |

**2.2.** **Scope** of application: what population and timeframe will your analysis/model be applied to or used for?

|  |
| --- |
| The population consists of cities with multi-family homes. The analysis will be done on data captured from 2011-2018 and used to predict 2019 data as our test set. The forecast will then be extrapolated to 2021 to see if the rental market has recovered from the effects of COVID-19 in cities. |

**3. How** do you translate the objective and scope in terms of data?

**3.1.** What **dataset**(s) do you plan to use? Initial description: source, granularity, number of observations, variables list…

|  |
| --- |
| We plan on using the following datasets:   * Zillow Rent Index (ZRI) for multi-family homes * Census data (ACS) * IRS tax data * Zillow Home Value Index (ZHVI) for multi-family homes |

**3.2.** What **data treatment and analysis** do you plan? Data Aggregation, target variable definition, tools, analysis/machine learning, ...

|  |
| --- |
| We plan on starting with a cluster analysis to determine different rental markets that act similar. We will also aggregate the various datasets, combining each on matching zipcodes. Our target variable will be the forecasted change in rent for each zipcode. We will apply multiple machine learning models to predict this, such as penalized regression, random forest, support vector regression, and time-series analysis. |

**4. Project plan**

Please submit a project plan proposal broken down by a few significant steps. Plan at least three meetings with your stakeholders.

- **Kick-off meeting**: schedule a 30 minute meeting before project declaration in order to approve the project proposal.

- **Milestone 1**: schedule 30 minutes with your stakeholders to present initial results and insights and to validate any assumptions or definitions needed to move forward.

- **Milestone 2**: schedule 30 minutes to go over the final results and proposed presentation before the final presentation in front of the whole team.

If you think additional discussions will be required, feel free to add secondary milestone(s) in your project plan.

**Project plan and schedule examples**

(create and use your own template)



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **September** | | | | | | | | | |
|  | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| **Kick off Mtg** |  |  |  |  |  |  |  |  |  |  |
| Project declaration |  |  |  |  |  |  |  |  |  |  |
| Data Wrangling |  |  |  |  |  |  |  |  |  |  |
| Cluster Analysis |  |  |  |  |  |  |  |  |  |  |
| **Project Mtg 1 & 2** |  |  |  |  |  |  |  |  |  |  |
| Basic Machine Learning |  |  |  |  |  |  |  |  |  |  |
| Machine Learning by Cluster |  |  |  |  |  |  |  |  |  |  |
| Analyze Results |  |  |  |  |  |  |  |  |  |  |
| Forecast vs Actual |  |  |  |  |  |  |  |  |  |  |
| **Project Mtg 3** |  |  |  |  |  |  |  |  |  |  |
| Make Presentation |  |  |  |  |  |  |  |  |  |  |