**Geospatial Analysis to predict rates of houses in Texas**



**Final Project: INFO 7390 Advances in Data Science**

**Karan Milan Thevar**

**Rahul Chandra**

**Pratik Patre**

**Overview:**

**The rate of a house depends on factors like the total area, type of house or how much is the development in the area. But, there are other aspect to this as well i.e. spatial components. Spatial components can be one of the most important factors in determining the price of a location. Some of these components include the fatality rate due to natural calamities, the crime rates or number of hospitals, police and other facilities like these in an area. These factors do play an important role for finding an appropriate house Rate.**

**Goal:**

**The goal of this project is to cover the following Use Cases:**

1. **To show a user 5 probable house listings as per his input values.**
2. **To predict the prices of these house listings after 10 years as per the historical data of the location. The property data that would be used for this prediction are as follows:**
3. **MarketHealthIndex**
4. **DaysOnMarket**
5. **ZHVI - zillow home value Index**
6. **MoM - month over month depreciation**
7. **YoY - year over year depreciation**
8. **ForecastYoYPctChange**
9. **DaysOnMarket**
10. **Price**
11. **Zillow Home value**
12. **To predict the house price after 10 years per geospatial data. The geospatial data that would be used to predict property price are as follows:**
13. **INJURIES\_INDIRECT**
14. **MAGNITUDE**
15. **CATEGORY**
16. **DAMAGE\_PROPERTY**
17. **MURD01**
18. **RAPE01**
19. **ROBB01**
20. **ASSA01**
21. **BURG01**
22. **LARC01**
23. **Police\_AccToZip**
24. **Hospital\_AccToZip**
25. **To find out the current house rate for State of Texas based on above all parameters.**
26. **To forecast the possibilities of calamities like tornado, storm, hurricane etc. within 10 years span and project them on the property listing so that the user gets a clear picture of how risky it would be to buy a property in that area.**
27. **To project areas feasible for builders to build property based on crime rates, fatality rate, number of hospitals and police stations nearby.**

**Datasets:**

**Storm Data:** <https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/>

**Police Data:** <http://www.50states.com/texas/police_departments.htm>

**Population Data:** <https://www.dshs.texas.gov/chs/popdat/downloads.shtm>

**Crime Data:**

****

**HospitalData:** <https://www.arcgis.com/home/webmap/viewer.html?useExisting=1&layers=b3813b2d3a054c378247bf32bcd8d203>

**Tools, Technologies and Packages:**

* **Azure Machine Learning Studio**
* **IBM Data Science**
* **ArcGIS Desktop**
* **ArcGIS Online**
* **Hadoop MapReduce**
* **SparkR or PySpark**
* **Jupyter notebook-Python**
* **RStudio**
* **Docker**
* **Luigi**
* **Amazon S3**

**Proposed Approach:**

**The following would be the plan of action:**

1. **Data wrangling and Preprocessing**

**Data Download**: Our first task is to programmatically download the data from

<https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/> - For Storm Data

<http://www.50states.com/texas/police_departments.htm> - For Police Data

1. **Feature Engineering, Missing data analysis and key Insights**

Handling missing data for creating clean data. Preforming variable selection needed to predict the house rate.

1. **Exploratory Data analysis**

We will perform exploratory data analysis to determine different summary matrices for an area.

Creating tableau dashboard of the summaries.

Using **ArcGIS** for performing Hot-spot analysis and Empirical Bayesian kriging for analysis in R.

Using **R ArcBridge** to get data from ArcGIS into R for analysis.

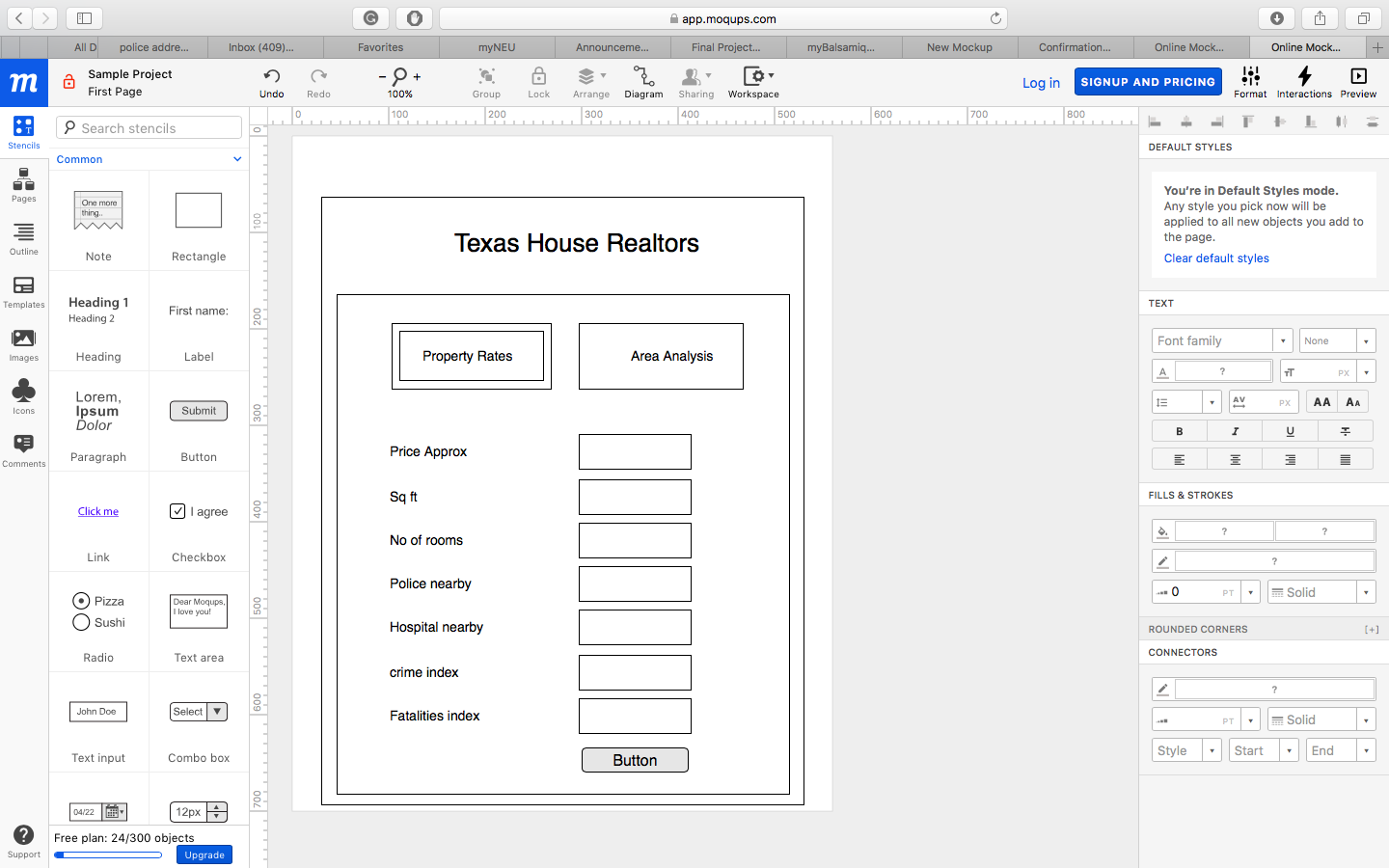
1. **Building and evaluating models**

The next step is to build models based on the data processed in the previous part to determine the property listings depending on geographical and property historical data.

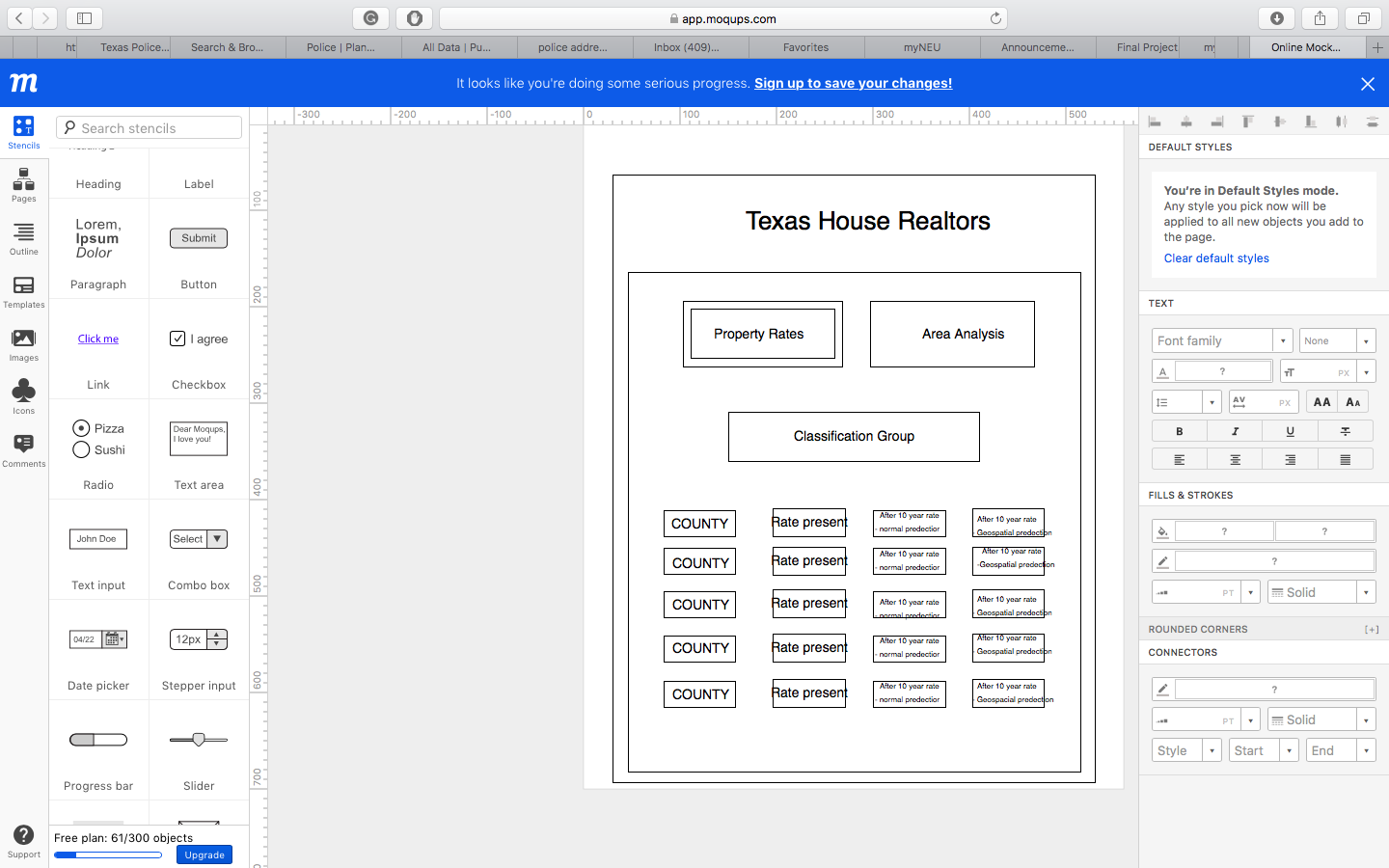
These models will be trained using the historical data. And based on this we will perform Classification, Clustering and Prediction.

**UI Design Plan:**

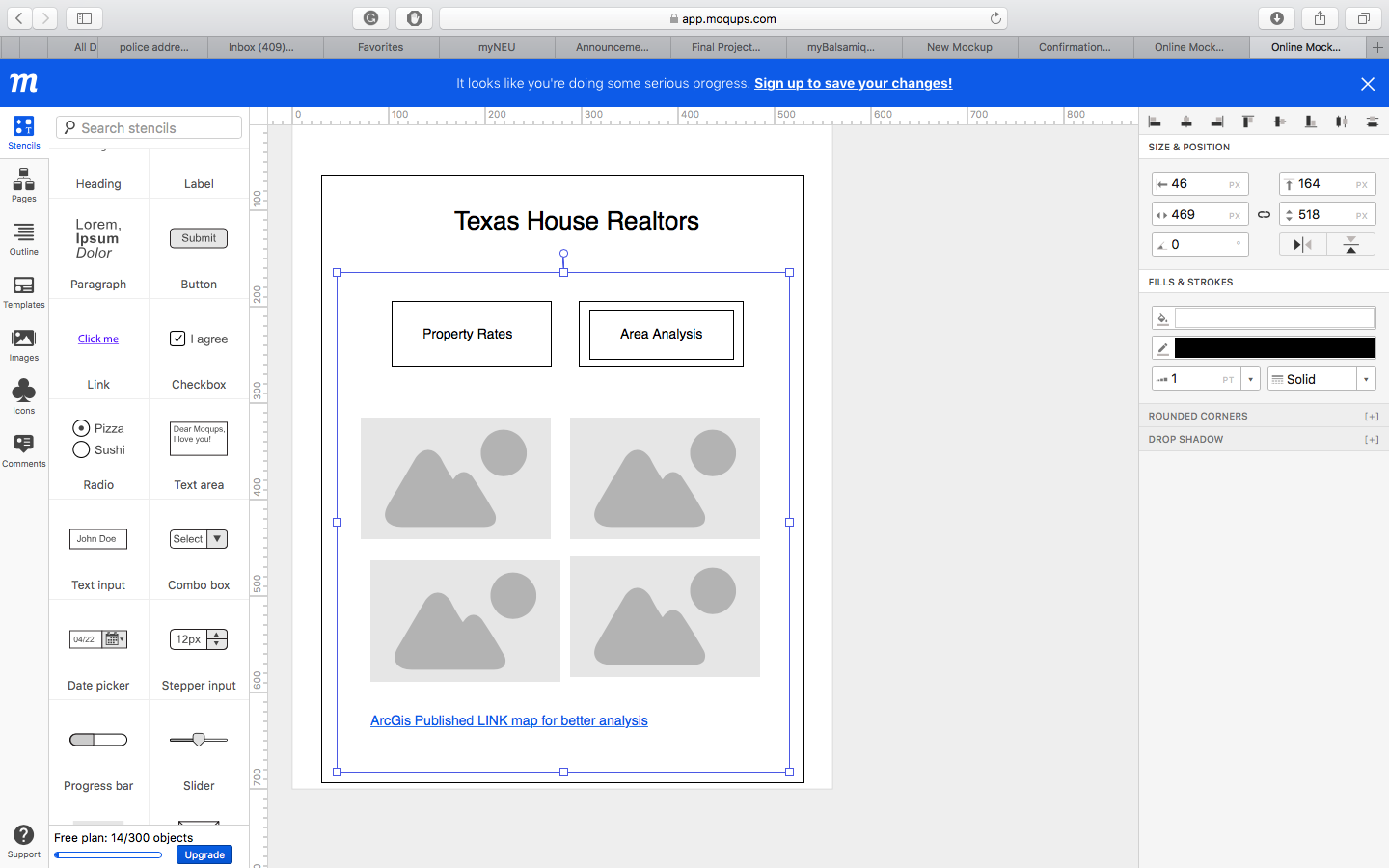
UI for User Input



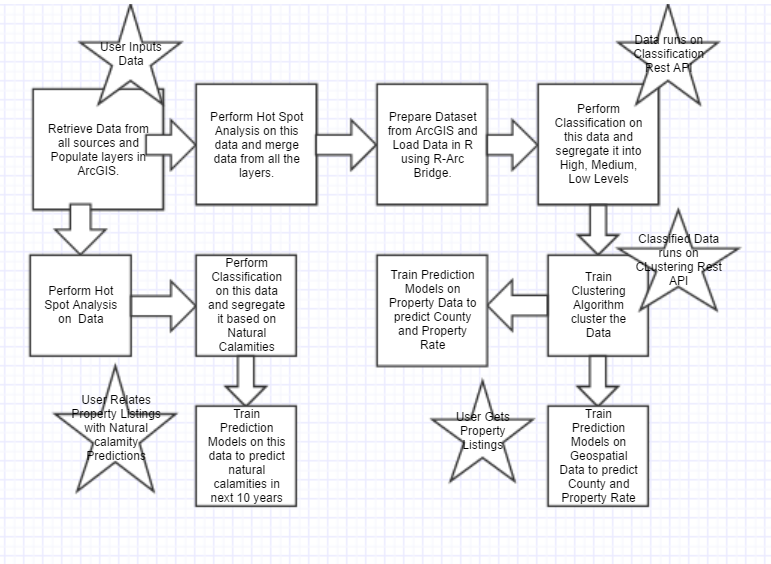
UI for Rates of Property Listing



UI for Tableau and ArcGIS Analysis



**Project Plan:**

****

**References:**

* <http://iireporter.com/4-business-trends-to-watch-in-the-insurance-industry-for-2017/>
* <http://desktop.arcgis.com/en/arcmap/10.3/main/get-started/arcgis-tutorials.htm>