

Market Maven

Daniel Bae

Keith Dowd

Interactive & analytics-based
real estate market research tool

Max Smoller

Ruoyi Chen

Introduction

Current tools are focused on searching individual real estate property listings and provide limited insights about real estate markets. Their data about markets is minimal and there is no capability to compare markets nor identify similar markets. This leaves a **large gap between supply and demand** for macro-level market data, which provides the main motivation for the creation of *Market Maven*. Our tool fills this niche by aggregating a diverse set of metrics, providing comparison capabilities, and **empowering home buyers, sellers, and investors** to discover new insights within a single, user-friendly web application.

Data

Key metrics are pulled from a combination of sources, including multiple Census API calls. Data for **3,142** counties, each with over **70** attributes, were wrangled and transformed using Python. These data, which were composed of **~618K** observations, were strategically condensed into a **46MB** SQLite database, as keeping this database small is crucial for a responsive, lag-free user web experience. The most recently available data from each source are displayed within *Market Maven*.

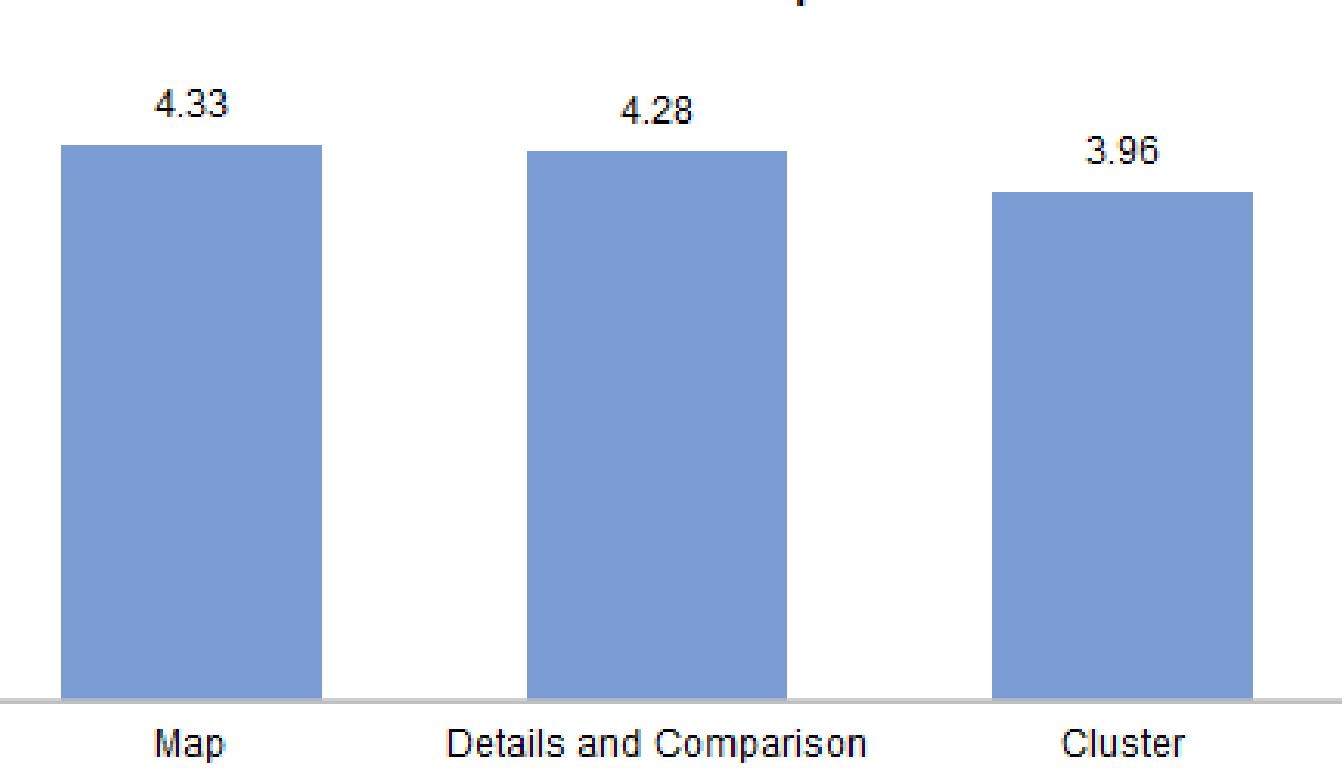
Experiments

Data was collected from two groups that each took a general market survey and a *Market Maven* survey in opposite orders. **F and t tests** were performed in order to determine whether seeing *Market Maven* first influenced their answers to both surveys. The mean, median, and coefficient of variation were calculated from the results in order to **assess sentiments**. Analysis on the **coefficients** of each of the features used in the k-means model were also conducted to evaluate our data and model.

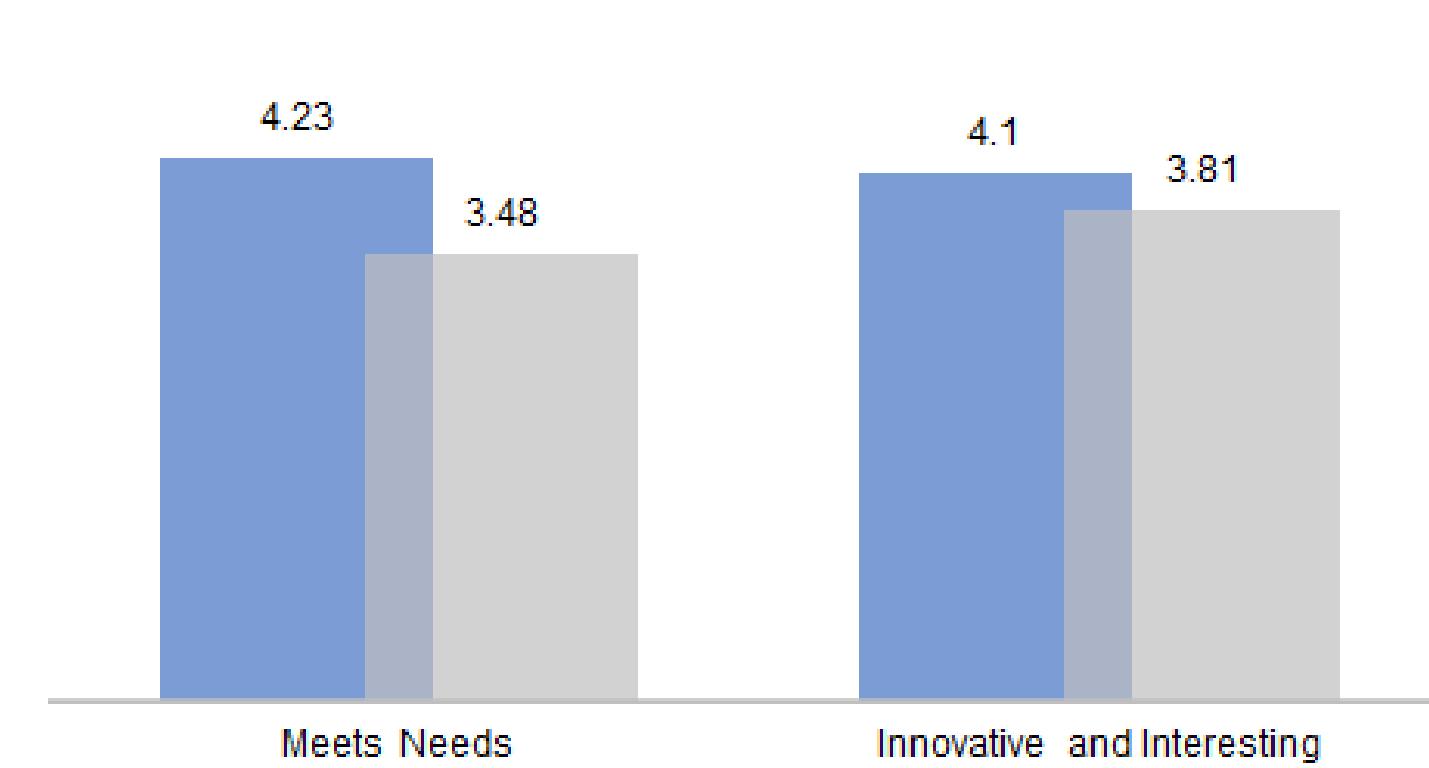
Results

The overall t test two-tail p-values for our surveys were 0.38 and 0.92. These results indicated that there was **no statistically significant difference** between the two groups. Our principal components analysis of the coefficients revealed that the first 18 of 32 components explained **99%** of variance. We could have improved how much of this information we captured in our k-means model but the set of features collected were informative, diverse, and uncorrelated. As shown below, *Market Maven* is more **innovative, interesting, and meets a need**, with some room for improvement (e.g. our clustering feature).

Feature Reception

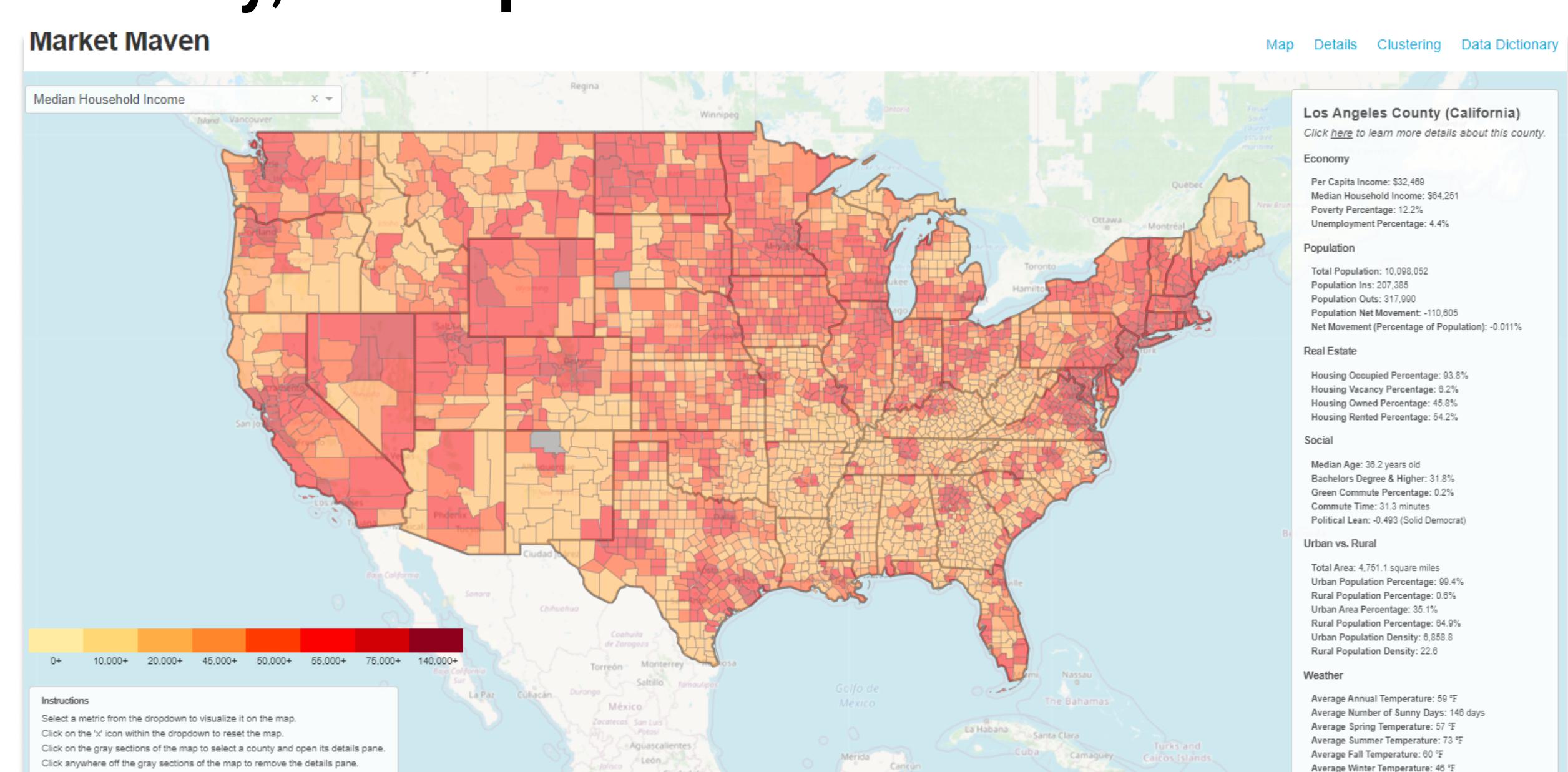


Overall User Sentiment

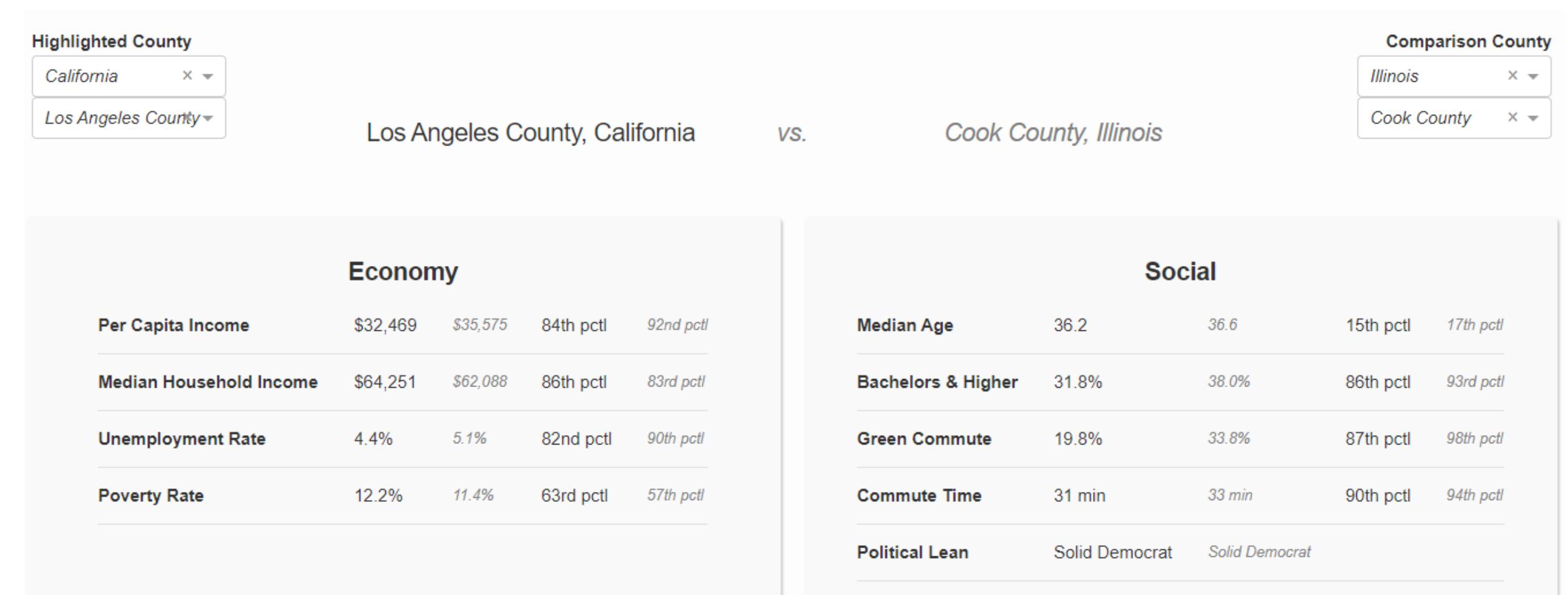


Our Approach

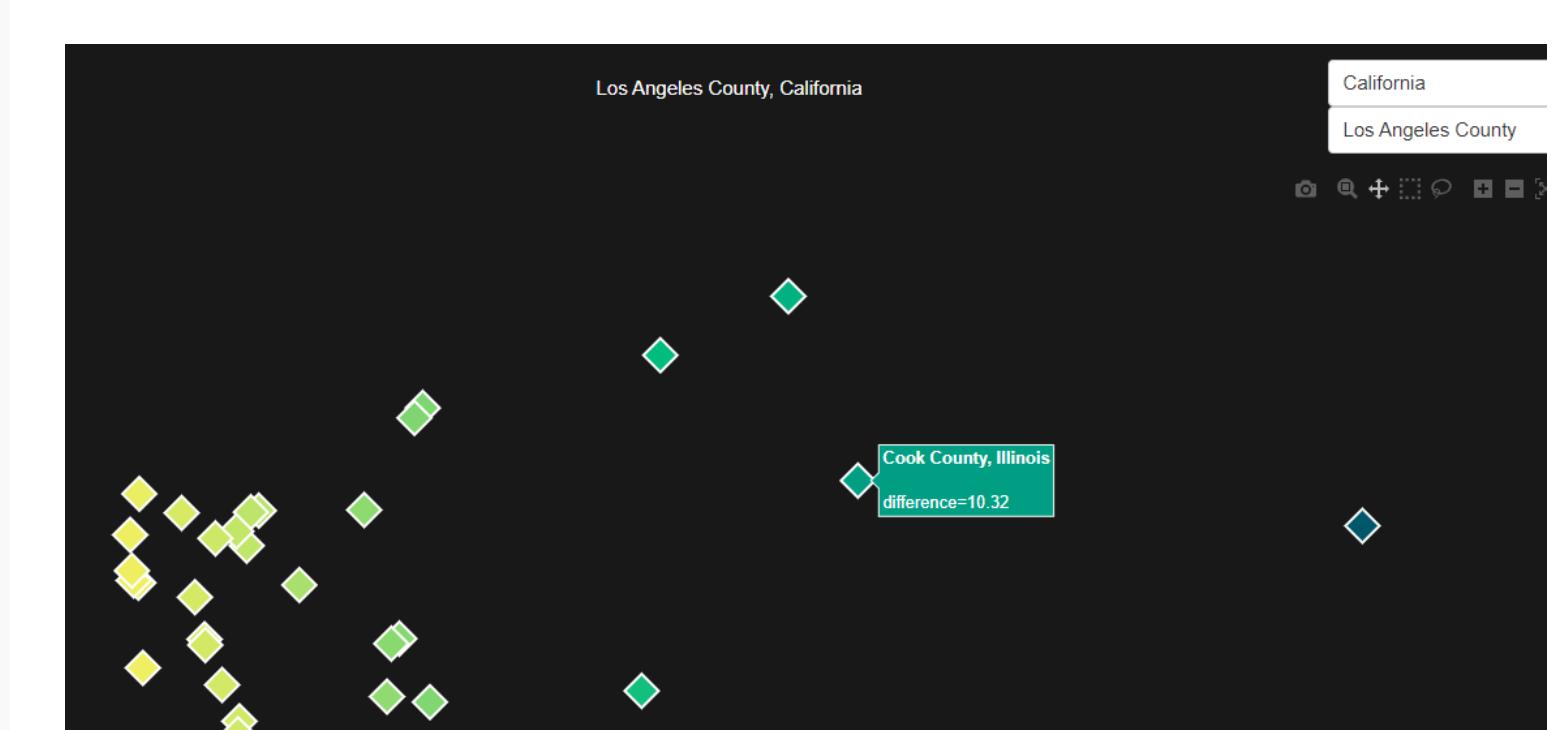
Market Maven is a Dash web application implemented in Python which differentiates itself by providing a suite of metrics not available in other tools, as well as choropleth maps, a market details and comparison feature, and the capability to identify similar markets. These features interact recursively to drive user **engagement, data discovery, and exploration**.



Utilizing Leaflet and geoJSON, the **map feature** allows the user to view and explore macro-level trends across the U.S. Users can select and visualize a variety of metrics, as well as select individual markets to view a convenient snapshot of data in the right-hand pane.



With the **market details and comparison feature**, a wide range of metrics are provided and are grouped into logical categories, such as economy and social. Users are able to select a county to compare with, and the metrics and percentiles for the comparison county are displayed side-by-side. A list of the top-5 similar counties derived from the clustering feature are also displayed.



The **clustering feature** utilizes PCA and a k-means model (Scikit-learn) to generate 60 clusters of similar counties. A Plotly visualization allows users to inspect the clusters by interacting with colored diamonds, which represent counties. Users can compare the similarity of nearby counties by hovering over the county and reviewing its similarity (difference) score.