**Project Hidden Figures**

**Project Problem and Hypothesis**

Advertising segmentations are becoming less effective and overused. New solutions in-market are trying to leverage social media, but only a limited population post information on social media and is not a representation of the larger population.

By leveraging location/movement data, we can understand a customers journey, their visitation to Points of Interest (POIs) and frequency to derive more accurate personas, brand affinities, and shift in lifestyle changes. This can then be stratified to a larger audience/universe for scale based on similar demographics and activity.

This project will consist of multiple models with the overall outcome being a prediction of a type of persona that a user falls into. A number of the models will leverage advanced analytics as well as some cognitive APIs from IBM to complete the overall project.

Once completed, this model will be deployed into a larger environment that contains a data set of location data and hosted on a cloud platform to run the analytics. These segmentations will feed several marketing engines across an enterprise and also be released out to a number of programmatic ad tech platforms.

**Datasets**

**Location Data** – Sample set of location data provided with hashed IDs that contain the following:

* IDFA (Advertising ID)
* Platform (Type of Device)
* Timestamp
* Lat/Long
* Point of Interest (Brand/Name)
* Sub Category (Category of POI)
* Accuracy Score (Quality of the ping)

**Twitter Data** – Use pacakage like Selenium to scrape twitter data back on each brand to run through IBM Personality Insights

**Alchemy API** – Use News API to extract Brand content and assign categories/concepts to Brands. Full documentation found here: <https://www.ibm.com/watson/developercloud/doc/discovery/index.html>

**Personality Insights** – 52 characteristics around needs and values that will help define a user’s persona. Full documentation here: <https://www.ibm.com/watson/developercloud/doc/personality-insights/getting-started.html>

**Domain knowledge**

I’ve had some experience working in this area and understanding the need of being able to classify profiles based on location data for marketing efforts. There is a real need to unlock new segments outside the typical DMA (marketing segments) that are currently used in the market today. I’ve had the opportunity to work with a number of these services listed in the data sets section, but have not put together the code to actually run them (have operated the demos).

There is some additional research required around potentially creating personality buckets/tagging using NLP and Spacy. This would simplify some of the clustering needed to move from Personality Insights to personas.

Others have ventured into this space, but no real public benchmarks available for this type of project.

**Project Concerns**

I’m somewhat concerned on being able to complete all the steps required to get to a final output. There is a bit of a concern also on getting my hands on the hashed ID data since most of the resources are currently taken up/deployed on a project.

The Brand evaluation to construct a user’s persona may not fully represent a user since Brand image being portrayed on Twitter is positioned based on the type of image the Brand would like to have, not necessarily the one that they do have or the reaction the customers of that Brand are having.

Cost of the model being wrong is that the personas/segments created are not accurate and clients lose faith in enterprise to provide relevant insight. The opposite is also true for being right, the market would respond favorably to current segments not available in market and gain a competitive advantage…result is $$$.

**Outcomes**

The expected outcome will be rows of IDs with columns that represent the personas in which the user is predicted to fit. In order for this project to be considered a success, a best practice/learning out of this will need to be taken away and applied to the business. Additionally, if a few new personas are created, it will be considered a major success. Finally, if we are able to take this code and implement/automated it as a hosted solution then we can run through the entire enterprise of IDs and attach segments/profiles to each (moonshot goal).