Barings Data Science Project Implementation Document

**Implementation Initiative**: Real Estate Supply Mapping (RESM)

**Sponsor(s):** FirstName LastName

**EDA Started:** 1/10/2019

**Proposed Implementation Start:** 09/15/19

Project Objective

The primary objective is to optimize the workflow of Real Estate research analysts by programmatically reconciling unique records from the DATAPROVIDER1 and DataProvider2 datasets into a master list of properties under construction.

The value-add achieved is a reduction on the time spent the research analysts allocate to manually generating these scrubbed lists of reconciled, unique records and free up resources for the Real Estate team to perform other functions.

Delivery Strategy: Intermediate & Long Term

The functionality delivered in this project is anticipated to ultimately sit on a centralized platform inclusive of access to the functionality from another actively developing Data Science project (CALA) and potentially already existing functionality in internally developed tools the Real Estate team currently utilizes.

The implementation of RESM individually, which is the intent of this document, is *not* inclusive of the larger effort to construct this centralized application to access these other functionalities. Should the need to develop functionalities that would require a more robust GUI arise from the clients, the functionality will be well documented and Strategic Development will be involved in a separately documented effort to drive the implementation of that kind of framework.

DS-External Cross-Functional Resources

**Data Services Resource** – Acquire DataProvider2 data as FTP from user-initiated drop into a specific shared folder; Crawl the DATAPROVIDER1 website to acquire update files of actively under construction records.

**PMO Resource** – Monitor and organize collaborative efforts across the rest of the cross-functional implementation team

**Data Engineering** – Advises & implements the data flow related components of the solution

**Data Architecture** – Help with deployment of API to interact with GUI & advise on deployment of web-hosted GUI

\*\* As mentioned above, **Strategic Development will be integral** to the effort to deliver the ultimate platform inclusive of the currently proposed UI functionality into a larger, more stable platform. Their availability to contribute to this effort will directly impact the delivery timeline.

Scope

Functions in Scope

* Batch pull of DATAPROVIDER1 Breaking Ground data from the DATAPROVIDER1 website
* Pipe DATAPROVIDER1 and DataProvider2 data into SQL server (likely Internal Platform)
* Create necessary tables, views, and stored procedures needed to facilitate GUI functionality
* Identify and remove any duplications in the DataProvider2 records themselves
* Recommend reconciliations between DataProvider2 and DATAPROVIDER1 data in ambiguous situations using the records’ address, bldg. name, year built, and coordinate data
* Allow users to indicate a DATAPROVIDER1 record is new (and not recommendations referenced above) to add it to the reconciled (“scrubbed”) list of properties
* Provide a GUI that will allow users to reconcile the data by identifying correct matches or indicating a record is new

Functions Out of Scope

* Incorporation of datasets outside of DATAPROVIDER1 and DataProvider2
* Reconciliation functionality for properties outside the US
* Additional QA measures, such as scraping news sites to further reconcile DataProvider2’s accuracy

Assumptions

* Between the DataProvider2 and DATAPROVIDER1 data, the sponsors have identified DataProvider2 as the baseline for accuracy, as the updates are more frequent
* The Real Estate team will facilitate the impact of analyst workflow enhancement by coordinating procedural changes to leverage the functionality provided from this tool

Constraints:

* The usability and functionality Barings Data Science can provide is limited to the conditions communicated in the DataProvider2 and DATAPROVIDER1 usage licenses
* Storage of geolocational data (namely coordinates) is prohibited in Google API
* The intermediate delivery solution may involve hosting the Dash application developed in EDA to make the results usable relatively quickly. Dash is not a the most robust framework, but given that we anticipate an integrated UI for RESM and related functionality, this strategy is motivated by the need to deliver the functionality in the interim before the centralized platform is ready.

Dependencies:

* The business partners in the Real Estate team will be responsible for understanding and communicating relevant contract terms in the DATAPROVIDER1 and DataProvider2 usage licenses to members of cross-functional project team. This includes individual employee access and data distribution constraints.
* Availability of resources across the necessary cross-functional teams will define the implementation timeline
* The usability of the tool will be partly determined by the effectiveness and reliability of the automated data ingestion orchestrated by the Data Services team
* The matching algorithm relies on the availability & quality of coordinates retrieved from the Google API