Plunk Data Engineer

Technical Task

Background

Plunk has acquired a new data source from a 3rd party vendor which contains a propensity score that estimates how likely the homeowner is to take out a home equity line of credit (HELOC) on their property in the next 12 months. Marketing would like to use this score to target user acquisition campaigns, but the data needs to be integrated with the core property data set so that other attributes can be used in targeting, and to understand how well populated the propensity score is.

Data

The data set propensity_scores.csv contains a sample of 10,000 records from the third-party vendor, including address information, assessor parcel number (aka APN) and the propensity score.

The data set core_property_data.csv contains a sample of 10,000 records from the core property data set, including some basic property attributes, address information and assessor parcel number.

Task

- Create an ETL program to:
  - Ingest data sets into a SQL database
  - Match/blend the two data sources (the columns/methods to use are up to you)
  - Report a measure of the population of the propensity scores in the resulting dataset
- Create an API service which exposes a query endpoint for this data
  - Parameters:
    - zipcode: the Zip Code of interest
    - count: the number requested records
  - Response: the top `count` addresses and propensity scores as JSON objects, ordered by propensity (descending), including for each:
    - Address
    - Propensity score

Submission Requirements

- Create a public git repository to hold your source-code
- Include a README file describing how to set up and run solutions
- Solution should not assume any particular database is already installed

1100 Bellevue Way
Suite 8A-1
Bellevue, WA 98004
info@getplunk.com
www.getplunk.com