2021

Address Linkage Key

KSU CAPSTONE PROJECT SPRING 2021

KATHRYN GREER, SARITHA GUDALA, KATHLEEN MCELVEEN, BYRON SMITH

Project Plan

Project: Address Linkage Key Tool

Date: 03/28/2021

Overview

AnalyticsIQ is the project owner and is requesting a tool or set of tools that can be used to calculate the delivery point code (DPBC) of a given address.
The tool(s) will be developed using python and spark with version control through git. This will be an iterative project that will implement the USPS guidelines[1] for determining the DPBC of an address.
The first viable iteration of the tool(s) will be able to correctly determine the DPBC of an address that is properly formatted and does not contain special characters or additional address information such as apartment or suite numbers. Subsequent iterations will be able to correctly determine the DPBC of an address with additional address information, special characters, or improper formatting. Additional iterations should provide useful information for data analytics such as the ability to determine if an address in a dataset contains specific types of delivery points such as a high-rise building.

Project Team

Roles	Name	Major responsibilities	Contact (Email and/or Phone)
Project owners	Warren Smith	Guide and instruct project team	warrens@analytics-iq.com
	Jin Wang	Guide and instruct project team	jinw@analytics-iq.com
Team leader	Kat McElveen	Scheduling, project documentation	Kmcelve2@students.kennesaw.edu, katmcelveen@gmail. 678-521-6870
Team members	Byron Smith	Technical Specialist	Bsmith513@students.kennesaw.edu, 404-644-6515
	Kat Greer	Technical Specialist	kgreer4@students.kennesaw.edu, 404-512-0468
	Saritha Gudala	Technical Writer: preparing required documents for presentation, reports.	sgudala@students.kennesaw.edu, 571-509-8207
Advisor / Instructor	Meng Han	Facilitate project progress; advise on project planning and management.	mhan9@kennesaw.edu

Project website:

https://sarithavikram.github.io/CapstoneProject_AddressLinkageKey/

Final Deliverables

1. A tool or set of tools which compute the delivery point code given:
    - A dataset with clean, properly formatted addresses
    - A dataset with ZIP and ZIP+4
    - A dataset with high-rise addresses
    - A dataset with P.O. Box, Rural Route(RR) and Highway Contract(HC) addresses
2. A tool or set of tools which compute the delivery point code given a dataset with improperly formatted addresses.
3. A tool or set of tools which determine if certain types of delivery points, such as high-rises, are included in the dataset.

Milestones

#1 - By 02/28/2021

- Analysis of tools
- Research on USPS formatting and processes

- Setup and installation of remote tools and development environment

#2 - By 3/28/2021

- Creation of a Minimally Viable Product.
- The tool or tools will determine the delivery point of a clean, properly formatted address, from a ZIP code and ZIP+4 appended.

#3 - By 4/11/2021

- Creation of an enhanced and optimized product.
- The tool or tools will determine the delivery point of an improperly formatted address.
- The tool or tools will determine if specific types of delivery points, such as high-rises, are included in the dataset.

Future milestone meetings date/time

- Wednesdays at 6:30 p.m. EST – internal team meetings
    - Milestone meetings:
        - Milestone #1: 02/26/2021 10:00 am to 11:30 am
        - Milestone #2: 03/26/2021 10:00 am to 11:30 am
        - Milestone #3: 04/09/2021 10:00 am to 11:30 am

Communication and Meeting Planning

All chat and meetings coordinated in Microsoft Teams.

- Weekly meetings with project owner AnalyticsIQ: Fridays 10:00 am to 11:00 am
- Weekly team meetings: Wednesdays 6:30 pm to 7:30 pm

Project Schedule and Tasks Planning

See the Gantt chart file attached.

Project Progress Report

Progress Made

- Project Team researched and tested regular expressions.
- Project Team completion of USPS defined primary delivery point barcode rules 1 through 13 (omission of rules 4 & 11 per project owner instruction).
- Project Team completed a minimally viable product (MVP) that calculates the DPC barcode of formatted address sets:
    - Git/TortoiseGit used to maintain code base and collaborate on DPC code development.
    - Unit test cases for each rule have been coded and run automatically for merge requests with the help of continuous integration (CI) in GitLab.

Next Steps

- Develop secondary address rules to enhance MVP. Emphasizing high-rise and apartment addresses.
- Expand primary address rules to account for PO Boxes, RR and HC addresses.
- Complete continuous delivery (CD) functionality to automatically deploy updates that pass unit tests to production (AWS).
- Updates to code to account for improperly formatted address source data.