Address Linkage Key

KSU CAPSTONE PROJECT SPRING 2021

KATHRYN GREER, SARITHA GUDALA, KATHLEEN MCELVEEN, BYRON SMITH

Project: Address Linkage Key with AnalyticsIQ
Date: 04/11/2021

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, following established guidelines from the United States Postal Service (USPS).

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.com, 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:
   o A dataset with clean, properly formatted addresses
   o A dataset with ZIP and ZIP+4
   o A dataset with high-rise addresses
   o A dataset with PO 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 - Milestone 3
Progress Made
- Completed a minimally viable product (MVP) that calculates the delivery point barcode of correctly formatted addresses according to USPS guidelines (initial tests indicate a roughly 90% correct rate).
- Created functions and test cases for USPS defined primary address rules (some rules omitted per project owner instruction).
- Expanded primary address rules to account for PO Boxes, RR and HC addresses.
- Created functions and test cases for USPS defined secondary address rules to enhance MVP (some rules omitted per project owner instruction).
- Created function and test case to calculate delivery point check digit.
- Learned how to run individual unit test cases, smoke tests, and regression testing at scale.
- GitLab used to maintain code base, keep track of tasks, and collaborate on code development.
- PySpark enabled Jupyter Notebok used for local testing and prototyping.

Next Steps
- Continue product enhancement and working on remaining secondary address rules in order to improve testing rate (goal: 95%).