2021

Address Linkage Key

KSU CAPSTONE PROJECT SPRING 2021

KATHRYN GREER, SARITHA GUDALA, KATHLEEN MCELVEEN, BYRON SMITH

PROJECT SITE:

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

Executive Summary

AnalyticsIQ, the project sponsor, is requesting a tool or set of tools that can be used to calculate the delivery point code (DPBC) of a given address. Having a tool to make this calculation will reduce reliance on third party tools.

The tool(s) will be developed using Python and Spark with version control through Git.

This will be an iterative project with the Minimally Viable Product (MVP) starting as a tool that will calculate the DPBC of a properly formatted address using the main USPS rule. Once the first iteration is complete and tested additional iterations will incorporate further USPS rules and the ability to determine the DPBC of improperly formatted addresses. An ultimate goal for the tool(s) is to determine if specific types of address, such as high-rise buildings, are included in a given dataset.

Project Presentation

Link for Project Formal Presentation:

Link for Project Short Presentation:

Introduction

Business and Project Background

AnalyticsIQ, the project sponsor, is a Marketing Data and Analytics company that works with big data from publicly available sources to allow their clients to deliver personalized experiences to consumers.

Address information is very useful in marketing, and there are existing software options that use the USPS guidelines to calculate information from addresses. AnalyticsIQ would like a tool or set of tools that would accomplish this task, determining the DPBC of a given address, so that address information can be used to help gain insights on the data they are analyzing and providing for clients. Creating a tool(s) would not only reduce reliance on third party software, but also allow for customization such as determining if certain address types, such as a high-rise building, are present in a given dataset.

Project Scope, Objectives, and Deliverables

This project was created to create a tool or set of tools that will determine the DPBC of a given address following the published USPS guidelines for determining DPBCs.

This is an iterative project that seeks to continually refine a minimally viable product (MVP). The first iteration will be to apply USPS rule number 1, which will determine the DPBC for most properly formatted addresses. Once the initial MVP is completed, additional iterations will compute the DPBC using additional USPS rules to process improperly formatted addresses or addresses that contain a secondary address component. Subsequent iterations will work to determine characteristics of addresses in a given dataset, such as high-rise buildings.

Analysis and Design

Technical Background

The tool will be designed using Python and Spark (PySpark) to allow for scalability to use the tool on Big Data datasets.

Version control and task assignments will be handled via Git. AnalyticsIQ created a GitLab site for this project where the code is stored. Task assignments are handled via a Kanban board on the project GitLab site. SSH is used to provide security when downloading or uploading code from the GitLab site to a local machine to work on the code.

Testing and prototyping will be done in Jupyter Notebook. AnalyticsIQ has provided a JupyterLab server instance for the team, and Docker containers will also be used to initialize PySpark-enabled Jupyter Notebooks locally.

Tool Design

Text goes here

Implementation

Summary of Tool Design

The tool will be designed using Python and Spark (PySpark) to allow for scalability to use the tool on Big Data datasets.

Version control and task assignments will be handled via Git. AnalyticsIQ created a GitLab site for this project where the code is stored. Task assignments are handled via a Kanban board on the project GitLab site. SSH is used to provide security when downloading or uploading code from the GitLab site to a local machine to work on the code.

Major Features of the Tool

The tool uses the published USPS guidelines to determine the DPBC of a given address.

Conclusion

Project Summary

Reflection on the project, deliverables met, lessons learned, etc.

Limitations and Future Direction

Future optimizations that would increase the usefulness of the tool(s)

References

CASS™ technical guide for CYCLE N. (n.d.). Retrieved February 28, 2021, from https://postalpro.usps.com/CASS/CASSTECH_N

USPS Publication 28. (2020, June). Retrieved 2021, from https://pe.usps.com/cpim/ftp/pubs/Pub28/pub28.pdf

Appendix

Attached Documentation/Information