Ziyao Tang

Professor Leonidas Kontothanassis

CDS DS210

1 May 2023

Report

To reach the goal that we are required to analyze a graph that is a collection of nodes connected by edges, I plan to study the data set of California Road Network's Edges. I used the data set published on the website of University of Utah by Lifeifei. The original data set of the road network is obtained from: Digital Chart of the World Server (http://www.maproom.psu.edu/dcw/) and the category points are obtained from: U.S. Geological Survey (http://www.usgs.gov/).

My project is divided into two parts. The first part is distance analysis and the second part is hub analysis. My goal is to find the shortest line between any two points and generate a graph.

For the hub analysis part, I compute the shortest path to any two points in the graph using Dijkstra's algorithm. I used python's folium module to generate a graph. By analyzing hubs, I discovered two patterns. The first one is that hub distribute along the main interstate highways in California, and the second one is that hub distribution concentrates in La and San Francisco.