Mini Challenge 1: Dataset - README

We provide you with two datasets. The first one contains microblog messages collected from various devices with GPS capabilities. These devices include laptop computers, handheld computers, and cellular phones. The second one contains map information for the entire metropolitan area. The map dataset contains a satellite PNG image with labeled highways, hospitals, important landmarks, and water bodies. We also provide supplemental tables for population statistics and observed weather data. Additional information is provided below.

**Microblog Messages**

The provided CSV file contains a number of microblog messages.

Attributes:

ID – personal identifier of the individual posting the message

Created\_at – date and time of the post

Location – latitude and longitude coordinates of the mobile device at the time of post

Text – the posted message

Example:

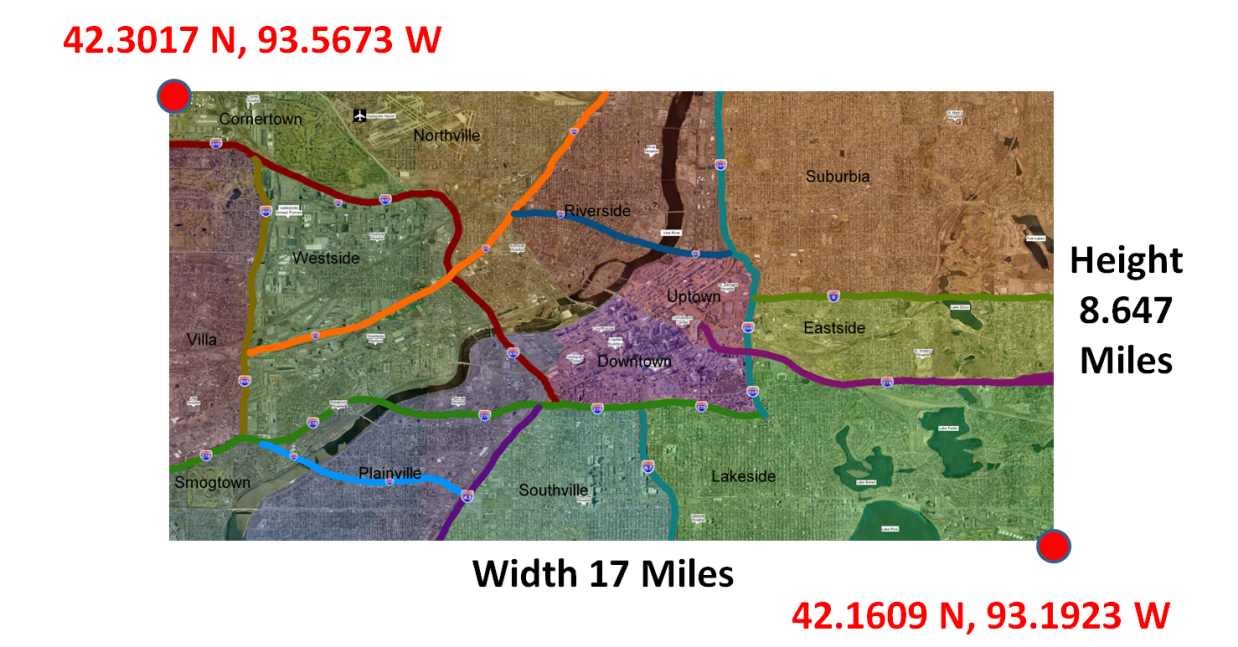
|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Created\_at** | **Location** | **Text** |
| 1001 | 12/22/2010 13:53 | 28.5220 -81.2838 | I think flu is on the way |
| 1002 | 12/22/2010 13:47 | 29.3783 48.0025 | Is up with a little boy that has the stomach flu! :( |
| 1003 | 7/25/2010 10:54 | 29.6009 -95.5645 | my flu is getting out of control =/ |
| 1004 | 12/22/2010 12:32 | 29.6009 -95.5645 | Man flu 'not a myth' say scientists |

Notice:

Every effort was made to remove offensive content from the collected microblog data. If any such content is found, please send the corresponding IDs to the challenge committee and consider those individual records removed from the dataset.

**Map**

The map is provided as a PNG image file.



Accompanying the image file, we provide two tables in CSV format.

POPULATION STATISTICS

Attributes:

Zone\_Name – the name of one of the 13 city zones within the metropolitan area

Population\_Density – the number of residents in the zone

Daytime\_Population – the estimated population in the zone due to commuting during work hours

OBSERVED WEATHER

Attributes:

Date – date of observed weather by weather station

Weather – weather conditions for a particular day

Average\_Wind\_Speed – measured in miles per hour

Wind\_Direction – the direction from which the wind is blowing or from which it originates

ADDITIONAL INFORMATION

1. Economy – The economy of Vastopolis is based on commerce, entertainment, finance, trucking services, shipping services, health care, and industry.
2. Water Supply - Residents and businesses get their drinking water by pumping water from nearby reservoirs or rivers. These distributed water systems are both public and privately owned.
3. Entertainment – Vastopolis has two stadiums (Vastopolis Dome and Westside Stadium) for sports, concerts, and other events. The various lakes and the Vast River, which flows south at a steady rate of three miles per hour, is used for water-based sports and recreation.
4. City Administration – Vastopolis has several locations of significance including a state courthouse, a capitol building, convention center, and a large airport.