Snowflakes - Data Discovery

1.Demographic Shift

[datahub](#)
Format: xlsx

Description: This is dataset includes information on change in the population of Hispanic or Non-Hispanic origin over period of 5 years.Other data includes Age,employment,median income of people living in Chicago.These details helps in understanding the attributes that will have impact on the lifestyle of people.

Attributes	Usefulness in the model
Year,Total population	Helps in comparing percentage of rise/decrease in population with Crime
Net International migration per 1,000 residents	Helps in inferring the reason for migration with crime,weather,political events
Educational attainment	Helpful in understanding crime by knowing the convicted qualification

2.MeteoBlue Weather Data

Yearly:
[meteoblue](#)

Monthly:
[meteoblue](#)

Format: csv

Description: The dataset provide information about the weather of City of Chicago. The data can be adjusted to get either the data for a month or year. Some of the data fields include Temperature, Precipitation and Snowfall mount etc.

Usefulness of data in Data Science:
- Also, it can be helpful in determining the possible anomalies in weather conditions such as the ones that may be due to global-scale climate change.
- By combining the data of present weather conditions and past patterns can help in making future weather predictions.

3.Chicago Community Area Data

datahub

Format: csv

Description: This dataset contains the data for 77 Chicago Community Areas. It includes data about Population, Employment, Land Use, Income, Demographics and Transportation for 2011-2015. The land use data can prove helpful in improving understanding of land use in terms of maintaining equitable land use for agriculture ,industries or houses. Another useful data field can be Mode of Travel To Work which can help in determining the most used mode for necessary improvements to the transportation system.

Data Fields for joining the datasets with other datasets: Chicago Community Area and Vacant Properties Data from City of Chicago data portal can combined on the community area name column. The common data field for two datasets are as follows:

	Chicago Community Area Data	**Vacant Properties Area(City of Chicago)**
Field Name	GEOG	Community Area

4. Divvy System Data:

divvybikes

Format: csv

Description: Here you'll find Divvy's trip data for public use. The data is available from Q3 2013 to 2017 Q2. We can analyze the movement of the residents of chicago and also analyze how often people use this particular mode of transportation. The most popular "hotspots" of the city can be identified.

The relevant attributes can be :
- "to_station_name"
- "from_station_name"
- "tripduration"
- "starttime"
- "stoptime"

We can use these details to infer people's movements and potentially utilize this data in various businesses like Restaurants, Shopping Centers, Tourist attractions and so on.

The **from_station_id | to_station_id** can be joined with dataset from Divvy Bicycle Station-Historical data **id** attribute to get detailed station information.

5. Data Made:

datamade
Trees in Chicago

Format: GDB

Description: Urban Tree Canopy Analysis from 2009 and shapefiles of trees in the City of Chicago. Data has become a 21st century design material. Architects, planners, engineers and citizens increasingly use data to understand urban issues and spark design innovation. This data can be utilized to check the land use, manage pollution levels in the future and make Chicago a more eco-friendly city.

6. Wunderground Weather Data

wunderground

Format: Web scrapable table

Description: This dataset includes information on weather history of Chicago over a year.The attributes such as Mean, Max and Min temperature gives us variance of temperature over a year and this data can be checked for any association rules with crime data over same interval.

7.Data about Schools

cps

Format: xlsx

Description: This contains the data (demographics, assessments, metrics, surveys) about the public schools in Chicago. With this data, we can analyze the performance of schools in different areas. This can be used to predict the growth of schools, population of different communities around those areas.

8. IDPH Population Projections For Chicago By Age And Sex 2010 To 2025:

data.illinois.gov

Format: csv

Description: This report presents projections of population from 2015 to 2025 by age and sex for Illinois, Chicago and Illinois counties produced for the CON Program. As actual future population trends are unknown, the projected numbers should not be considered a precise prediction of the future population; rather, these projections, calculated under a specific set of assumptions, indicate the levels of population that would result if the assumptions about each population component (births, deaths and net migration) hold true.

We can utilize this data to make sure that a particular business has a long term plan to cater to all possible demographics that may be present in Chicago in the future.

9. Business Lists:
chicagobusiness

Format: xlsx

Description: This website is a combination of several smaller datasets. These datasets provide information about the Accounting, Advertising, Banking, Real Estate, Corporate, Hospitals, Hotels, Insurance, Law firms and so on. All these datasets can be widely used in a variety of use cases and can have a large impact on any data science project.

Using this data we can project business performance in the future and link these datasets to practically any other data retrieved from data.cityofchicago.org, census.gov, datausa.io.

10. Zip Code Latitude Longitude City State County
gaslampmedia.com

Format : csv

Description: This dataset helps in connecting different datasets which has at least any of the following attributes: Zip Code, Latitude, Longitude, City, State or County of Chicago.The data consists of entire US and on filtering Chicago area will make the dataset useful.

11. Census
datahub

Format: xlsx

Description: This dataset includes information on different ages of people living in various regions of Chicago. Data of Age range from 0 to 85+ in both male and female categories.

Attributes	Usefulness in the model
Geography,Race-Not Hispanic or Latino, Black or African American alone	Helps in comparing percentage of rise/decrease in Crime for each race
Age group of male and female in a region	Can be helpful in inferring if weather has impact

	Census	Chicago Community Area Data
Field to join on	GEOGNAME	GEOG

12. Chicago Crime 2014-2015

chicagocrime

Format: csv

Description: This dataset contains the Crime Rate categories of not getting arrested in Wards, Districts and communities of Chicago.One way in which this data can be used is for determining the dangerous areas/ communities in Chicago as these are the areas where criminal commits crime but don't get arrested.

This dataset's **Type** field can be combined with the City of Chicago's Crimes -2001 to present dataset's **Primary Type** on the basis of type of crime.