

School Immunizations

You are given dataset of immunization(vaccination) status of kindergarten students in California in schools from year 2016-2019.The annual assessment was conducted to measure the immunization coverage amongst the students entering kindergarten.

Dataset Description

Data Column	Description	Type
Year	Year of immunization	String
School Code	Unique code of each school	Int
Country	Country name	String
School Sector	public or private	String
City	City name	String
School name	School name	String
Reported	Immunizations report was conducted or not(Y /N)	String
Enrollment	Total number of enrollees	Int
Category	Types of Immunizations (Polio,PBE,MMR etc)	String
Count	Total number of enrollees vaccinated	Int
Percent	Percentage of enrollees vaccinated	Int

Deliverable

The project will have three deliverables. Keep on adding the deliverables to the same document. This way, final report will be comprehensive report.

1. Exploratory Data Analysis (EDA) – 40% weightage
2. Clustering and Frequent Pattern Mining -40% weightage
3. Recommendations Based on Insights from Data/Final Report -20% weightage

1. Exploratory Data Analysis (EDA)

This deliverable will be focused towards getting familiar with the data. For that you will have to get your hands dirty with the data. This will consist of initial data exploration, visualizations etc. This could include but not limited to the following points:

- a. Data Preprocessing (Identifying duplicate data and removing it, finding missing values and replacing them by an appropriate strategy, normalizing the data)
- b. Calculating and visualization of summary statistics (consider boxplots, bar graphs)
- c. Finding correlation between attributes.
- d. Finding out dependence between categorical and numerical attributes.
- e. Annual Immunizations by school sector
- f. Immunization category by percentage of immunization
- g. Immunization category by school sector (think of many more)

2. Clustering and Frequent Pattern Mining

This deliverable will be more focused toward features selection and engineering for data mining tasks. Based on selected set of features you will perform specific tasks that can extract useful insights from data. **We are just listing directions to think, you can come-up with something else by which more meaningful information from data can be deduced.**

- a. Perform frequent pattern mining to identify patterns between immunization categories (Given a row, this task aims to find the rules that enable us to predict the occurrence of an immunization based on the occurrence of another immunization).
- b. Perform cluster analysis on the data.
- c. Report your findings.

3. Final Report

For this deliverable, focus on following points. Try to apply all the concepts learned during the course. You would not be told which techniques to apply but rather what objectives to achieve. You should be able to identify the approaches and techniques. Your conclusions, recommendations, reasoning, and findings should be supported by the relevant graphs/visualizations considering the energy domain. You need to identify what meaningful information can be extracted from this data.

Dataset Link

https://drive.google.com/file/d/1d6TUrIA2Aj6dHOBxYGQPq1A_H0iJX489/view?usp=sharing