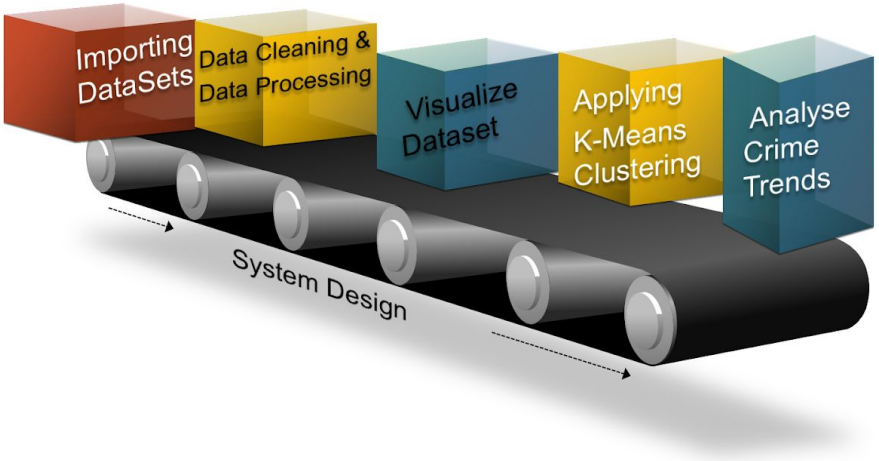
System Design



The System design is as follows:

1. **Importing data set:** The dataset used in the project is downloaded from kaggle, link of which is mentioned below. https://www.kaggle.com/currie32/crimes-in-chicago#Chicago_Crimes_2012_to_2017.csv
2. **Data cleaning:** The dataset had many null values and columns which were not necessary hence they were removed and a clean and good dataset was obtained
3. **Data exploration:** To have a good understanding of data, several plots like arrest vs non arrest, word cloud for frequent crimes, frequency plot of crime by types were plotted so that we can have some insights from the data.
4. **Data Modeling:** The nature of crimes varies in the dataset therefore we have used a clustering model. K-means is the most suitable algorithm for this analysis since it is easy to build clusters using k-means. We have made clusters with the help of data containing location, type and also with time and IUCR.
5. **Analysing Model:** The clusters obtained after applying k-means help us to draw a number of outcomes. Also the frequency plots shows the percentage of arrest, primary crime type etc