**.Report.**

Analysis of All India Pin code directory with contact details along with Latitude and longitude Dataset

All India Pin Code Directory contains all the pin-code list across India with geocodes and other relevant information like Office Name, Office Type, Delivery Status, Division, Region, Circle, Taluk, Districts, States, Telephones, Related Sub office and Related Head office etc. Postal Index Number (PIN) or PIN Code is a 6digit code of Post Office numbering used by India Post. The PIN was introduced on August 15, 1972. There are 9 PIN regions in the country. The first 8 are geographical regions and the digit 9 is reserved for the Army Postal Service. The first digit indicates one of the regions. The first 2 digits together indicate the sub region or one of the postal circles. The first 3 digits together indicate a sorting / revenue district. The last 3 digits refer to the delivery Post Office.

**Dataset**

The dataset used for this project was found data.gov.in. The basic idea of analysing the All India Pin code directory with contact details along with Latitude and longitude Dataset is to get fair idea about the factors affecting the establishment of different types of office names at different pin codes in all over the India.

This kind of analysis can be done using the data, by studying the factors such as Office Name, Office Type, Delivery Status, Division, Region, Circle, Taluk, Districts, States, Telephones, Related Sub office and Related Head office etc.

|  |  |  |  |
| --- | --- | --- | --- |

**Tools and Libraries**

* Python
* Jupyter Notebook
* Pandas
* Numpy
* Seaborn
* Matplotlib
* Plotly & Cufflinks

**Data Description**

This dataset contains 154797 rows and the following 15 columns:

1. Office Name
2. Pin Code
3. Office Type
4. Delivery Status
5. Division name
6. Region name
7. Circle name
8. Taluk
9. Districts name
10. States name
11. Telephones
12. Related Sub office
13. Related Head office
14. Longitude
15. Latitude

**Data cleaning**

I made the following changes

Deleted the columns like Taluk, Districts name, States name, Telephones, Related Sub office, Related Head office, Longitude, Latitude as they have more null values and were not important for my analysis

**EDA**

I looked at the different-different trends of the data and below are few highlights of the analysis.

* Imported the libraries which are essential for this analysis.
* Imported the all India pin code directory dataset for further analysis.
* Plotted a heatmap to find null values from the dataset.
* Found out the null value columns and deleted them for further analysis.
* Also deleted unwanted columns which are not important for this kind of analysis.
* Plotted a histogram of column called pin code to show the frequency.
* Plotted a box plot to show the max and min value of column called pin code.
* Plotted a bar plot to show top 30 office names.
* Plotted a bar graph to show top 30 value counts of column called office names.
* Plotted a bar plot to show top 10 delivery status
* Plotted a scatter plot to show the reference between two columns called pin code and circle name.
* Plotted a histogram to show a group of all columns.