Problem Statement

Analyze the Aadhar data for various demographic parameters, the analysis involves basic data preparation, processing and understanding. Process the data and generate insights.

Data Description:

Date: Registration Date
Registrar: Name of Registrar Office
Private Agency: Name of private agency working for registration of Aadhar Cards
State: Name of State
District: Name of District
Sub District: Name of Sub District
Pin Code: Postal code
Gender: Gender
Age: Age group
Aadhar Generated: Total no of aadhar card generated on a particular day
Rejected: Total no of enrollment rejected on a particular day
Mobile No: Count of residents, provided the mobile no at the time of enrollment.
Email Id: Count of residents, provided the email id at the time of enrollment.

You can understand the data dictionary better by the following example: A row with data - 20150420, Allahabad Bank, A-Onerealtors Pvt Ltd, Uttar Pradesh, Ambedkar Nagar, Akbarpur, 224155, F, 15, 5, 0, 0, 4 indicates that

- On 20 Apr 2015 (date), for A-Onerealtors Pvt Ltd (private_agency) registered with Allahabad Bank (registrar) at PIN code 224155, Akbarpur (sub_district), Ambedkar Nagar (district), Uttar Pradesh (state)
- Among the group of women aged 15
- There were 5 Aadhar numbers generated and 0 were rejected
- Out of the 5 that applied, none had an email ID and 4 had mobile numbers

Data Download Link:

https://drive.google.com/file/d/0B6rxRECSt4WdVkMzOTE2ODVWcG8/view?usp=sharing

Problem Statements

KPI-1

1. View/result of the top 25 rows from each individual store

KPI-2

1. Find the count and names of registrars in the table.
2. Find the number of states, districts in each state and sub-districts in each district.
3. Find the number of males and females in each state from the table.
4. Find out the names of private agencies for each state

KPI-3

1. Find top 3 states generating most number of Aadhaar cards?
2. Find top 3 private agencies generating the most number of Aadhar cards?
3. Find the number of residents providing email, mobile number? (Hint: consider non-zero values.)
4. Find top 3 districts where enrolment numbers are maximum?
5. Find the no. of Aadhaar cards generated in each state?

KPI-4

1. Write a command to see the correlation between "age" and "mobile_number"? (Hint: Consider the percentage of people who have provided the mobile number out of the total applicants)
2. Find the number of unique pincodes in the data?
3. Find the number of Aadhaar registrations rejected in Uttar Pradesh and Maharashtra?

KPI-5

1. The top 3 states where the percentage of Aadhaar cards being generated for males is the highest.
2. In each of these 3 states, identify the top 3 districts where the percentage of Aadhaar cards being rejected for females is the highest.
3. The top 3 states where the percentage of Aadhaar cards being generated for females is the highest.
4. In each of these 3 states, identify the top 3 districts where the percentage of Aadhaar cards being rejected for males is the highest.
5. The summary of the acceptance percentage of all the Aadhaar cards applications by bucketing the age group into 10 buckets.