Aadhaar Use Case Analysis using HIVE

Use case of analyzing a dataset from ***Aadhaar*** - a unique identity issued to all resident Indians. The issuing authority - UIDAI provides a catalog of downloadable datasets collected at the national level. our dataset of interest is "Enrollments processed in detail".

**Problem Statement:**

* CSV file contains following columns: Registrar, Enrollment Agency, State, District, Sub District, Pin Code, Gender, Age, Aadhaar Generated, Enrollment Rejected, Residents providing email, Residents providing mobile number
* Count the number of Identities generated in each state
* Count the number of Identities generated by each Enrollment Agency
* Top 10 districts with maximum identities generated for both Male and Female

Questions of Interest or other Analytics that has to be drawn from the Aadhaar data

1. For how many states the Aadhaar data exists?
2. What is the % of Aadhaar's being approved per state? Are there states where there is a high percentage of rejection?
3. What is the percentage of people applying for Aadhaar per state? Are there some states where the percentage of people applying for Aadhaar are low? (This would need census information per state) For the states with high rate of rejection can we narrow down to a district or an agency?
4. What is the distribution of Aadhar cards per age?