# TOP 5 STATE WISE LOAN ELIGIBILITY ANALYSIS WITH OPENREFINE & GCP

**Sai Dinesh Kumar Reddy Addula**
**Vamshi Kodali**
**Mohana Srinitha Shaga**

# OBJECTIVE

To strategically establish a new bank branch in one of the top 5 GDP states in India, using data-driven insights based on:

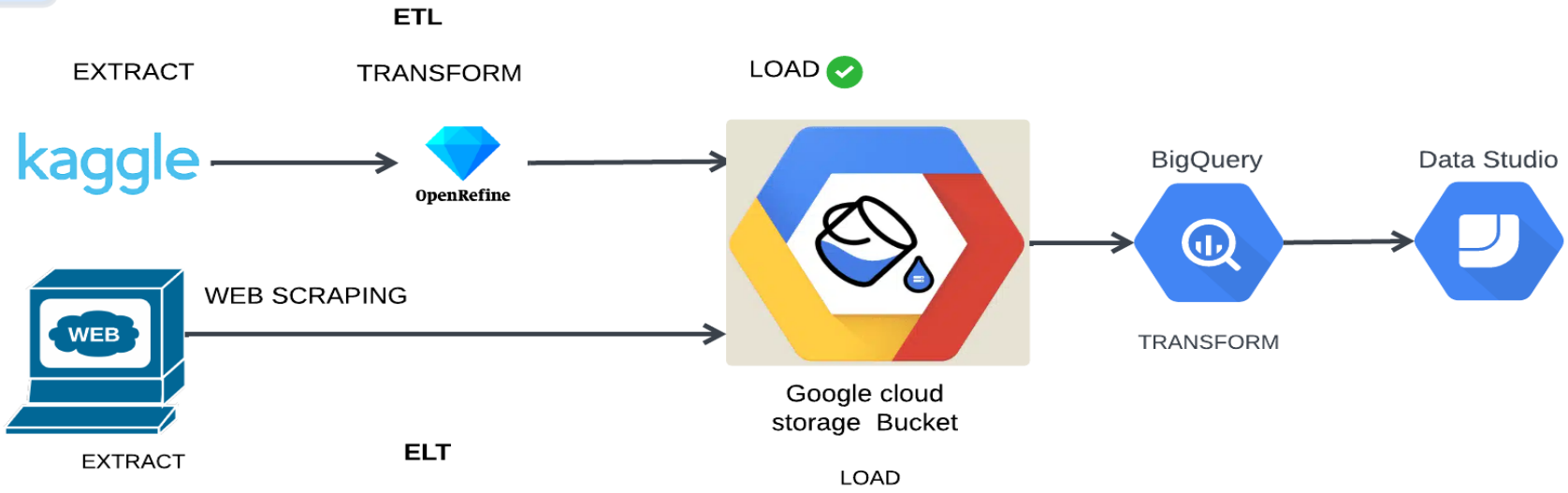
- The total number of loan applications received from each state.
- The average income of applicants in these states.
- The number of applicants with over 4 years of work experience.
- The number of applicants who have resided at their current address for more than 2 years.
- The total number of applicants with age greater than 25.

Datasets

Kaggle

Web Browser

# FLOW CHART



# ✓ DATA COLLECTION

## Kaggle

### Applicant Details For Loan Approve

34 | New Notebook | D

Data Card | Code (17) | Discussion (1) | Suggestions (0)

| use_Ownership | Vehicle_Ownershi... | Occupation | Residence_City | Residence_State |
|---|---|---|---|---|
| d 92% |   true 30.3k 30% false 69.7k 70% | Physician 2% | 317 unique values | Uttar_Pradesh 11% |
| d 5% | | Statistician 2% | | Maharashtra 10% |
| (2831) 3% | | Other (95236) 95% | | Other (78587) 79% |
| ed | no | Psychologist | Jalandhar | Punjab |
| ed | no | Petroleum_Engineer | Bally | West_Bengal |
| ed | no | Drafter | Indore | Madhya_Pradesh |
| l | no | Chartered_Accountant | Kurnool[18] | Andhra_Pradesh |
| ed | no | Air_traffic_controller | Asansol | West_Bengal |
| ed | no | Biomedical_Engineer | Bongaigaon | Assam |
| ed | yes | Fashion_Designer | Katihar | Bihar |
| ed | yes | Magistrate | Sikar | Rajasthan |
| ed | no | Technology_specialist | Chinsurah | West_Bengal |

## Web

**List of Indian states and union territories by GSDP (2021–22)**[3][4]

| | State/union territory | GSDP (₹ millions) | GSDP (US$ billions) |
|---|---|---|---|
| 1 | Maharashtra * | 31,080,218.7 | 417.2 |
| 2 | Tamil Nadu * | 20,712,861.6 | 278.0 |
| 3 | Gujarat * | 19,770,664.6 | 265.4 |
| 4 | Uttar Pradesh * | 19,745,315.5 | 265.1 |
| 5 | Karnataka * | 19,627,254.2 | 263.5 |
| 6 | West Bengal * | 13,639,258.6 | 183.1 |
| 7 | Rajasthan * | 12,181,934.5 | 163.5 |
| 8 | Madhya Pradesh * | 11,361,371.9 | 152.5 |
| 9 | Andhra Pradesh * | 11,338,365.0 | 152.2 |
| 10 | Telangana * | 11,289,072.4 | 151.5 |
| 11 | Kerala * | 9,324,699.6 | 125.2 |
| 12 | Delhi * | 9,046,420.4 | 121.4 |
| 13 | Haryana * | 8,706,645.3 | 116.9 |
| 14 | Odisha * | 6,708,812.3 | 90.1 |
| 15 | Bihar * | 6,503,024.3 | 87.3 |
| 16 | Punjab * | 6,142,267.6 | 82.4 |
| 17 | Assam * | 4,126,118.7 | 55.4 |
| 18 | Chhattisgarh * | 4,064,158.2 | 54.6 |
| 19 | Jharkhand * | 3,588,625.9 | 48.1 |
| 20 | Uttarakhand * | 2,721,594.8 | 36.5 |
| 21 | Jammu and Kashmir * | 1,999,174.2 | 26.8 |

# ✓ DATA CLEANING: KAGGLE DATA PROCESSING

OpenRefine Mohana Srinitha Shaga Permalink

**100000 rows**

Show as: **rows** records Show: 5 10 25 50 100 500 **1000** rows

| All | | Residence_State | Loan_Default_Risk | Applicant_ID | Annual_ |
|---|---|---|---|---|---|
| | 1. | Punjab | 0 | 75722 | 9657655 |
| | 2. | West_Bengal | 0 | 80185 | 9259353 |
| | 3. | Madhya_Pradesh | 0 | 19865 | 1509721 |
| | 4. | Andhra_Pradesh | 1 | 76700 | 5867312 |
| | 5. | West_Bengal | 0 | 92992 | 7223191 |
| | 6. | Assam | 0 | 76435 | 1161425 |
| | 7. | Bihar | 1 | 84005 | 7059529 |
| | 8. | Rajasthan | 0 | 80918 | 555485 |
| | 9. | West_Bengal | 0 | 60768 | 5395461 |
| | 10. | Uttar_Pradesh | 1 | 50075 | 4772948 |

OpenRefine Mohana Srinitha Shaga Permalink

**Add column based on column Residence_State**

New column name: Residanc_State

On error: set to blank / store error / copy value from original column

Expression Language: General Refine Expression Language (GREL)

```
value.replace("_", " ")
```

No syntax error.

Preview | History | Starred | Help

| row | value | value.replace("_", " ") |
|---|---|---|
| 1. | Punjab | Punjab |
| 2. | West_Bengal | West Bengal |
| 3. | Madhya_Pradesh | Madhya Pradesh |
| 4. | Andhra_Pradesh | Andhra Pradesh |
| 5. | West_Bengal | West Bengal |
| 6. | Assam | Assam |

OpenRefine Mohana Srinitha Shaga Permalink

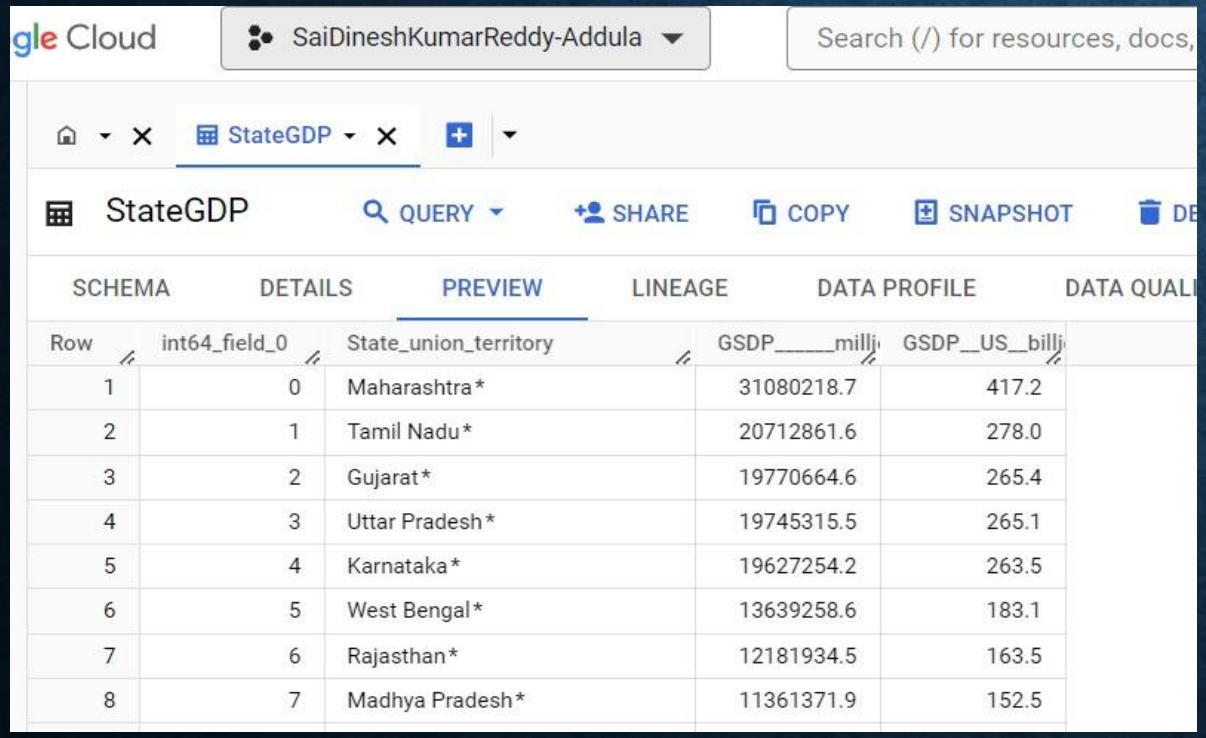
**87003 matching rows** (100000 total)

Show as: **rows** records Show: 5 10 25 50 100 500 **1000** rows

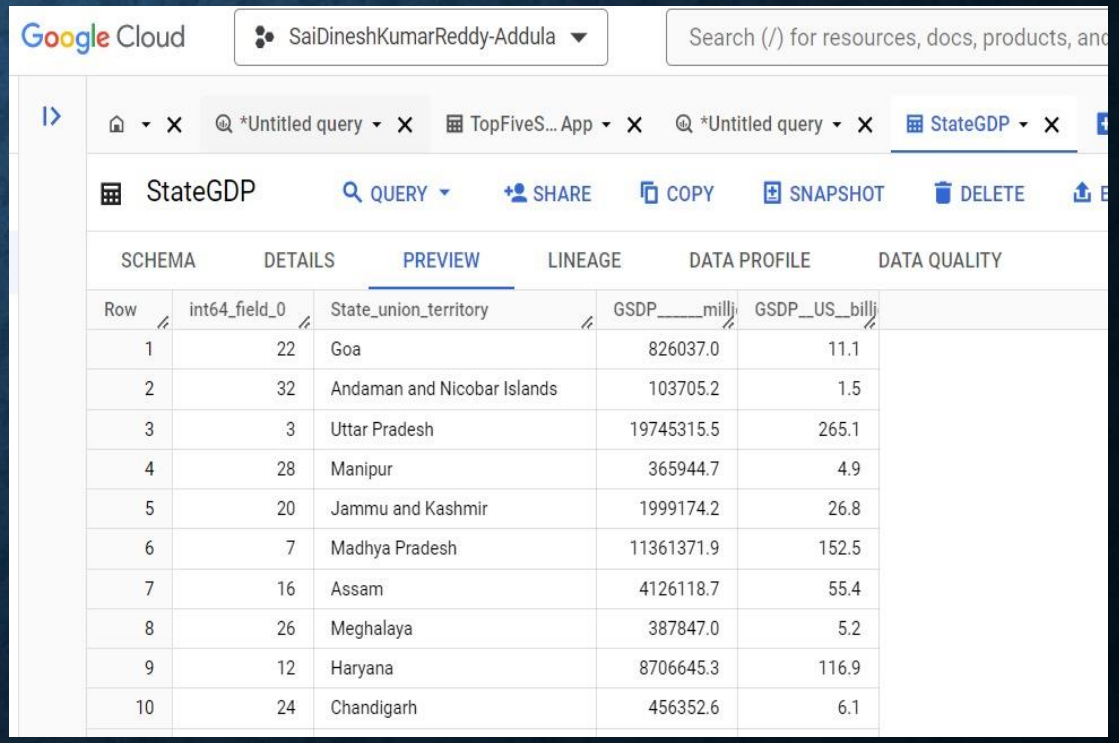
| All | | Residanc_State | Loan_Default_Risk | Applicant_ID | Annual_Incom |
|---|---|---|---|---|---|
| | 1. | Punjab | 0 | 75722 | 9657655 |
| | 2. | West Bengal | 0 | 80185 | 9259353 |
| | 3. | Madhya Pradesh | 0 | 19865 | 1509721 |
| | 5. | West Bengal | 0 | 92992 | 7223191 |
| | 6. | Assam | 0 | 76435 | 1161425 |
| | 8. | Rajasthan | 0 | 80918 | 555485 |
| | 9. | West Bengal | 0 | 60768 | 5395461 |
| | 11. | Bihar | 0 | 27702 | 5098928 |
| | 12. | Chandigarh | 0 | 42142 | 6259059 |
| | 13. | Andhra Pradesh | 0 | 45081 | 5888261 |
| | 15. | Uttar Pradesh | 0 | 20425 | 3140216 |

# ✓ WEB DATA PROCESSING

```
UPDATE ‘saidineshkumarreddy-addula.Group8Project.StateGDP’
SET State_union_territory = REPLACE(State_union_territory, ‘*’, ‘ ’) where State_union_territory LIKE ‘%’;
```



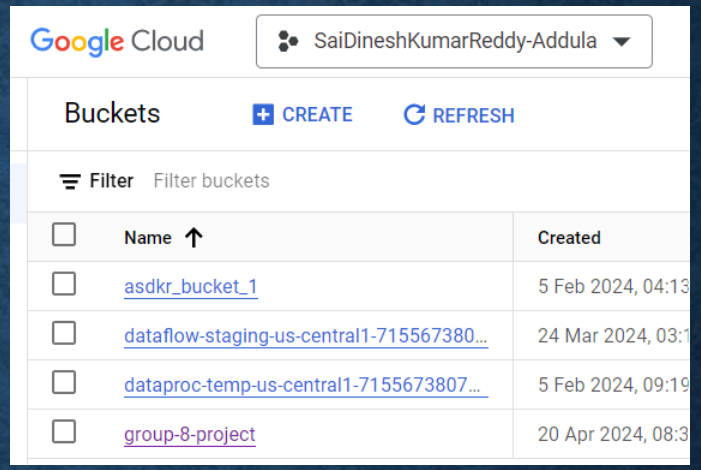

| Row | int64_field_0 | State_union_territory | GSDP_____milli | GSDP__US__billi |
|---|---|---|---|---|
| 1 | 0 | Maharashtra* | 31080218.7 | 417.2 |
| 2 | 1 | Tamil Nadu* | 20712861.6 | 278.0 |
| 3 | 2 | Gujarat* | 19770664.6 | 265.4 |
| 4 | 3 | Uttar Pradesh* | 19745315.5 | 265.1 |
| 5 | 4 | Karnataka* | 19627254.2 | 263.5 |
| 6 | 5 | West Bengal* | 13639258.6 | 183.1 |
| 7 | 6 | Rajasthan* | 12181934.5 | 163.5 |
| 8 | 7 | Madhya Pradesh* | 11361371.9 | 152.5 |



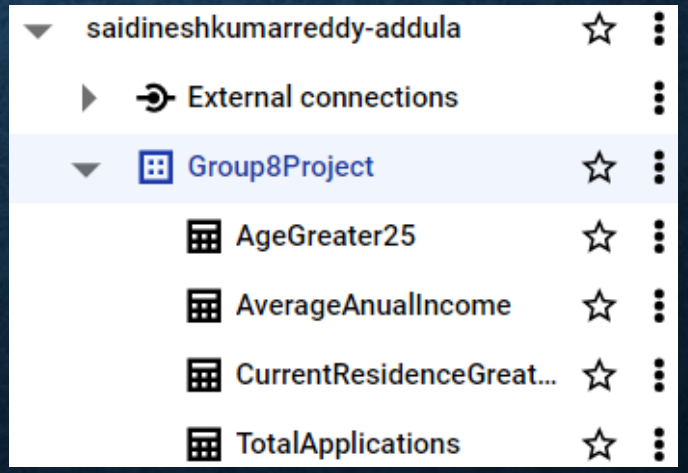

| Row | int64_field_0 | State_union_territory | GSDP_____milli | GSDP__US__billi |
|---|---|---|---|---|
| 1 | 22 | Goa | 826037.0 | 11.1 |
| 2 | 32 | Andaman and Nicobar Islands | 103705.2 | 1.5 |
| 3 | 3 | Uttar Pradesh | 19745315.5 | 265.1 |
| 4 | 28 | Manipur | 365944.7 | 4.9 |
| 5 | 20 | Jammu and Kashmir | 1999174.2 | 26.8 |
| 6 | 7 | Madhya Pradesh | 11361371.9 | 152.5 |
| 7 | 16 | Assam | 4126118.7 | 55.4 |
| 8 | 26 | Meghalaya | 387847.0 | 5.2 |
| 9 | 12 | Haryana | 8706645.3 | 116.9 |
| 10 | 24 | Chandigarh | 456352.6 | 6.1 |

# ✓ TYPES OF STORAGES USED IN GCP

## Cloud Storage(object storage)

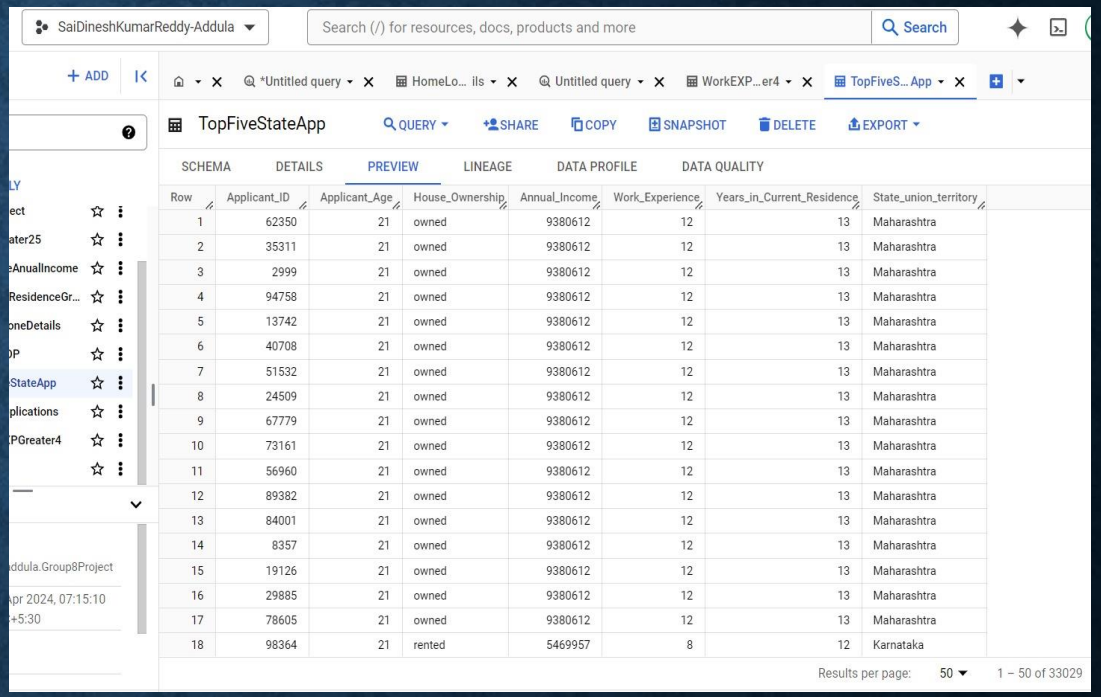




## BigQuery(data warehouse)

# ✓ TOTAL APPLICATIONS

```
CREATE TABLE
'saidineshkumarreddy_addula.Group8Project.TopFiveStateApp' AS
SELECT
    tb1.Applicant_ID,tb1.Applicant_Age,tb1.House_Ownership,
    tb1.Annual_Income, tb1.Work_Experience,
    tb1.Years_in_Current_Residence,
    tb2.State_union_territory
FROM
    (SELECT TRIM(State_union_territory) AS State_union_territory,
GSDP_USbillions
     FROM saidineshkumarreddy_addula.Group8Project.StateGDP
     ORDER BY GSDP_USbillions DESC
     LIMIT 5) AS tb2
INNER JOIN
    saidineshkumarreddy_addula.Group8Project.HomeLoneDetails AS tb1
ON
    tb1.Residence_State = tb2.State_union_territory;
```



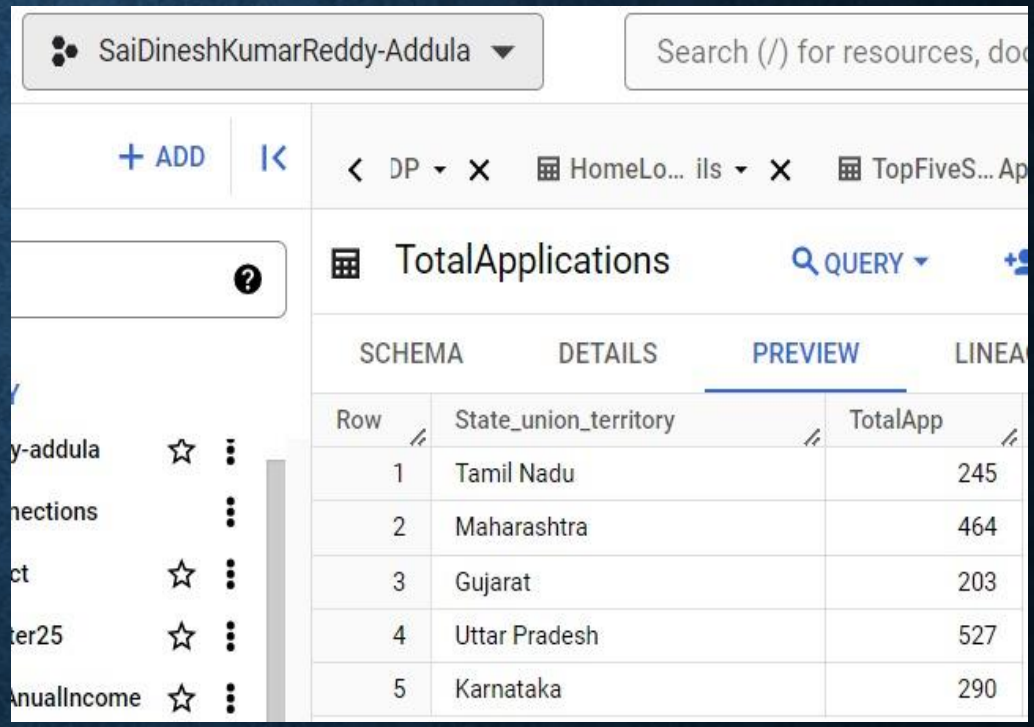

| Row | Applicant_ID | Applicant_Age | House_Ownership | Annual_Income | Work_Experience | Years_in_Current_Residence | State_union_territory |
|---|---|---|---|---|---|---|---|
| 1 | 62350 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 2 | 35311 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 3 | 2999 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 4 | 94758 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 5 | 13742 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 6 | 40708 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 7 | 51532 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 8 | 24509 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 9 | 67779 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 10 | 73161 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 11 | 56960 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 12 | 89382 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 13 | 84001 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 14 | 8357 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 15 | 19126 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 16 | 29885 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 17 | 78605 | 21 | owned | 9380612 | 12 | 13 | Maharashtra |
| 18 | 98364 | 21 | rented | 5469957 | 8 | 12 | Karnataka |

Results per page: 50 1 – 50 of 33029

- The query generates a new table called TopFiveStateApp, which combines home loan applicant data from the five states with the greatest GSDP in your database. It combines the HomeLoneDetails table with a subquery of the top five states by GSDP, using residence state to state territory matching.

# ✓ TOP FIVE STATE APPLICATIONS

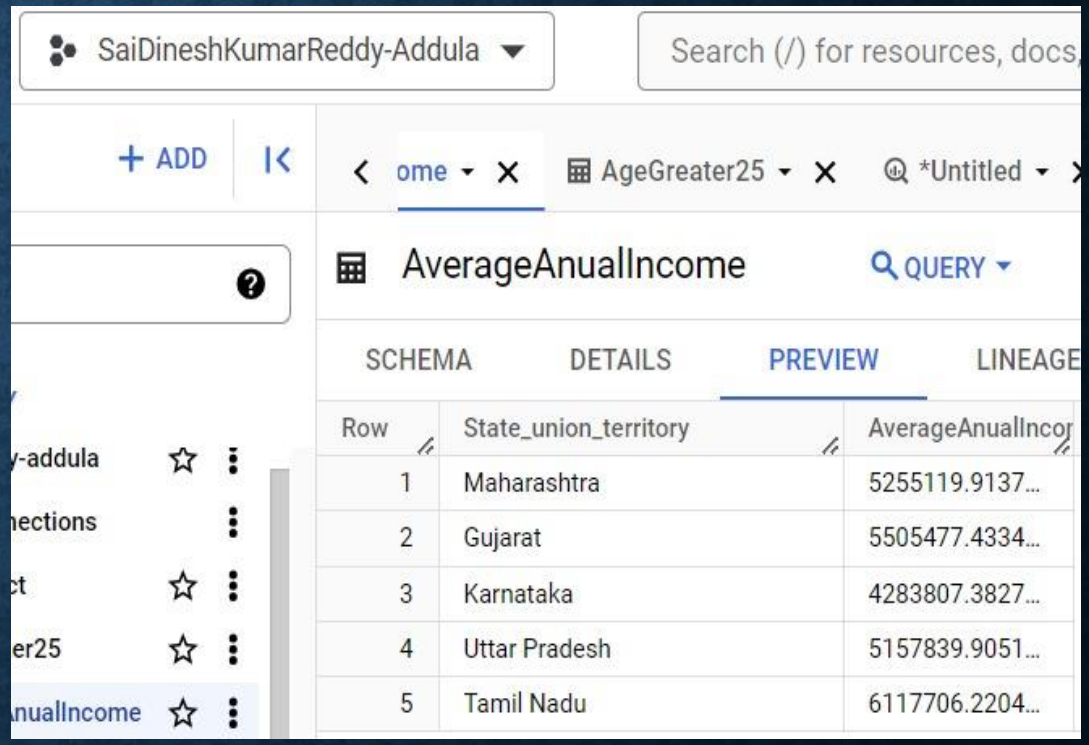
```
CREATE TABLE saidineshkumarreddy_addula.Group8Project.TotalApplications AS
SELECT
   State_union_territory,
   COUNT(*) AS TotalApp
FROM
   saidineshkumarreddy_addula.Group8Project.TopFiveStateApp
WHERE
   House_Ownership = 'owned'
GROUP BY
   State_union_territory;
```




- This generates a fresh table for storing the count of home loan applications with owned houses per state from the top five states based on GSDP.

# ✓ AVERAGE ANNUAL INCOME

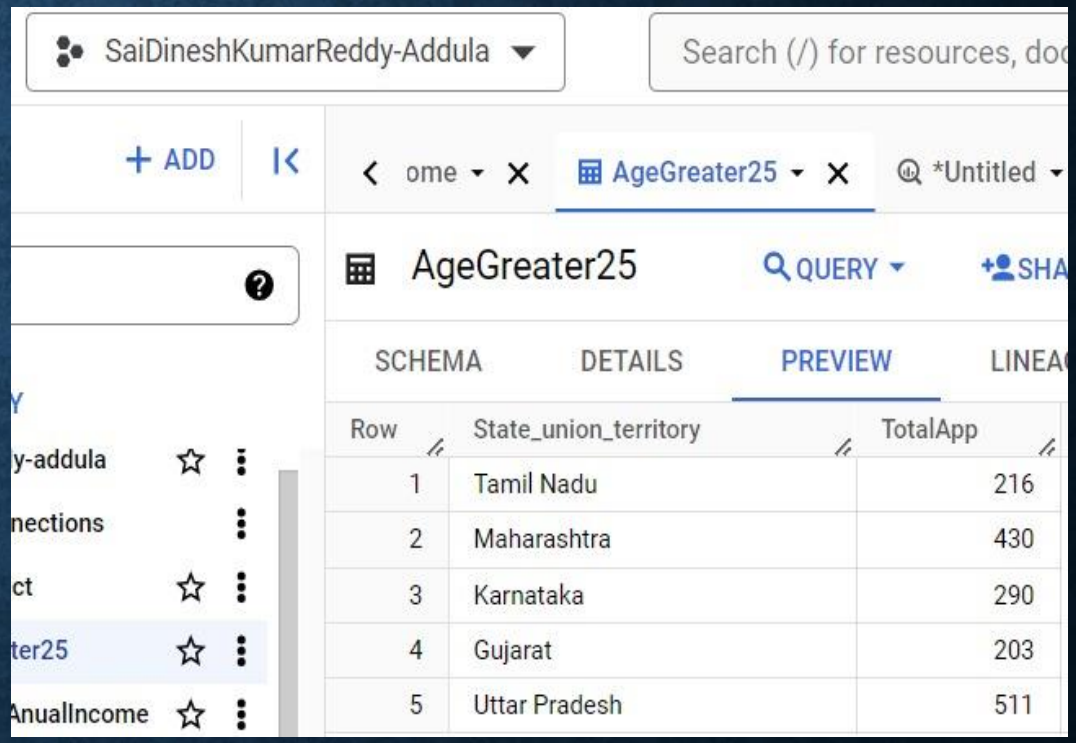
```
CREATE TABLE saidineshkumarreddy_addula.Group8Project.AverageAnnualIncome
AS
SELECT
   State_union_territory,
   AVG(Annual_Income) AS AvgAnnualIncome
FROM
   saidineshkumarreddy_addula.Group8Project.TopFiveStateApp
WHERE
   House_Ownership = 'owned'
GROUP BY
   State_union_territory;
```




- This generates a new table that stores the average yearly earnings of individuals who have applied for home loans and own homes in the top five states with the highest GDP.

# ✓ AGE GREATER THAN 25

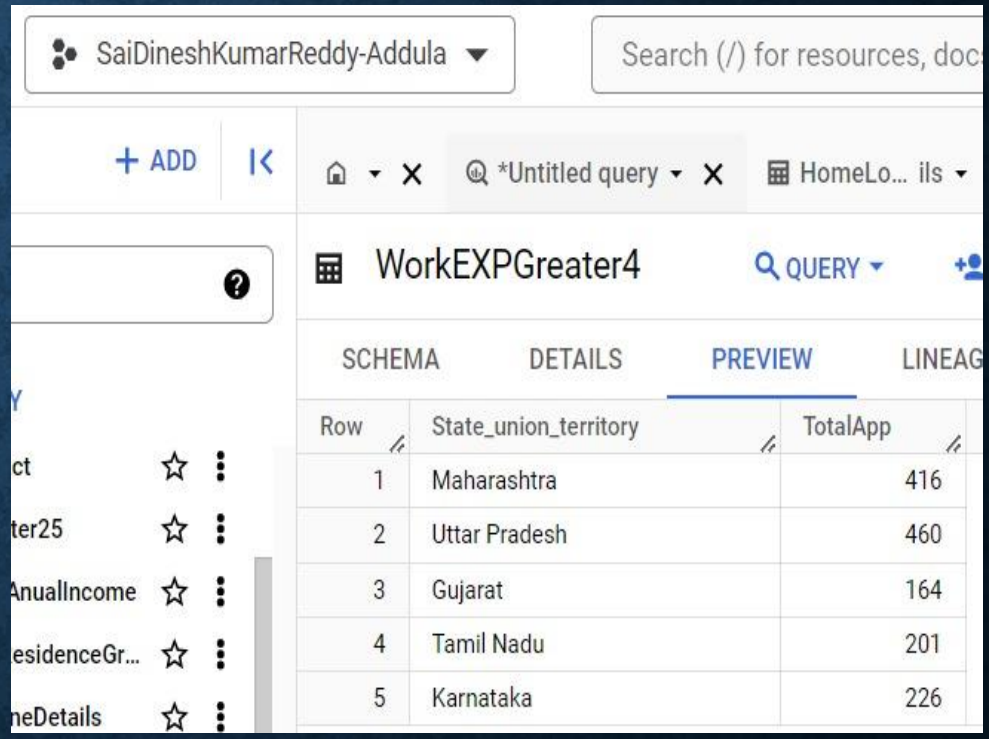
```
CREATE TABLE saidineshkumarreddy_addula.Group8Project.AgeGreater25 AS
SELECT
    State_union_territory,
    COUNT(*) AS TotalApp
FROM
 saidineshkumarreddy_addula.Group8Project.TopFiveStateApp
WHERE
    House_Ownership = 'owned' AND Applicant_Age >= 25
GROUP BY
    State_union_territory;
```




- This generates a table displaying the number of home applicants aged 25 or above, categorized by state.

# ✓ WORK EXPERIENCE GREATER THAN 4

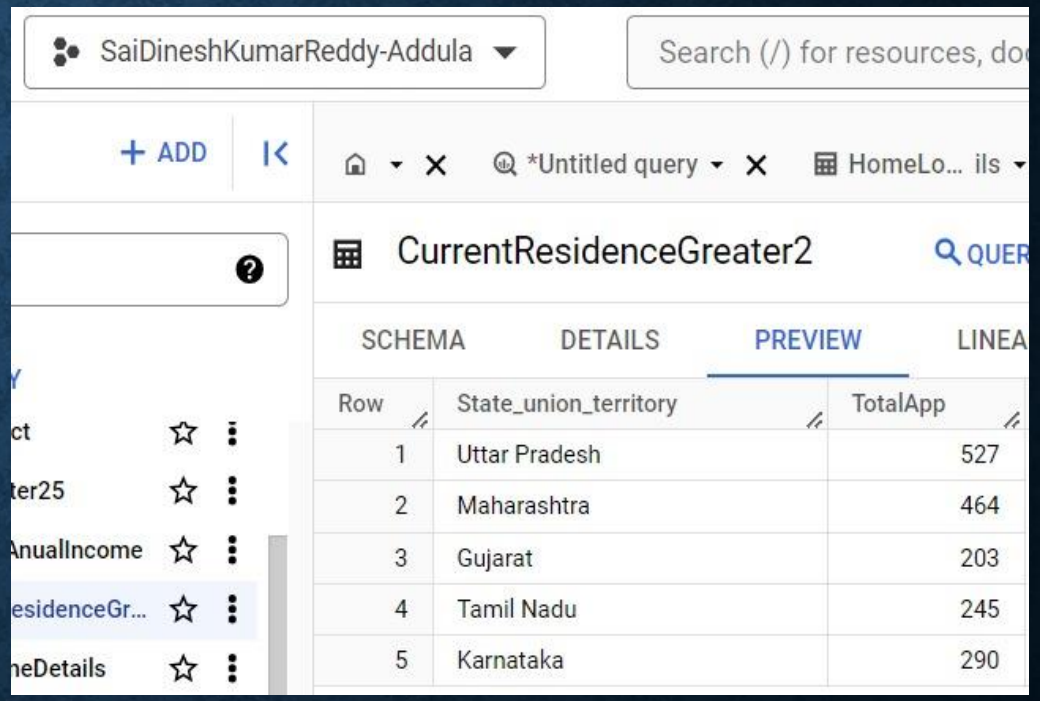
```
CREATE TABLE
saidineshkumarreddy_addula.Group8Project.WorkEXPGreater4
AS

SELECT

   State_union_territory,

   COUNT(*) AS TotalApp

FROM

   saidineshkumarreddy_addula.Group8Project.TopFiveStateApp

WHERE

   House_Ownership = 'owned'

   AND Work_Experience >= 4

GROUP BY

   State_union_territory;
```




- The query creates a new table listing each state or territory and the count of loan applications where the applicant owns a home and has 4 or more years of work experience.

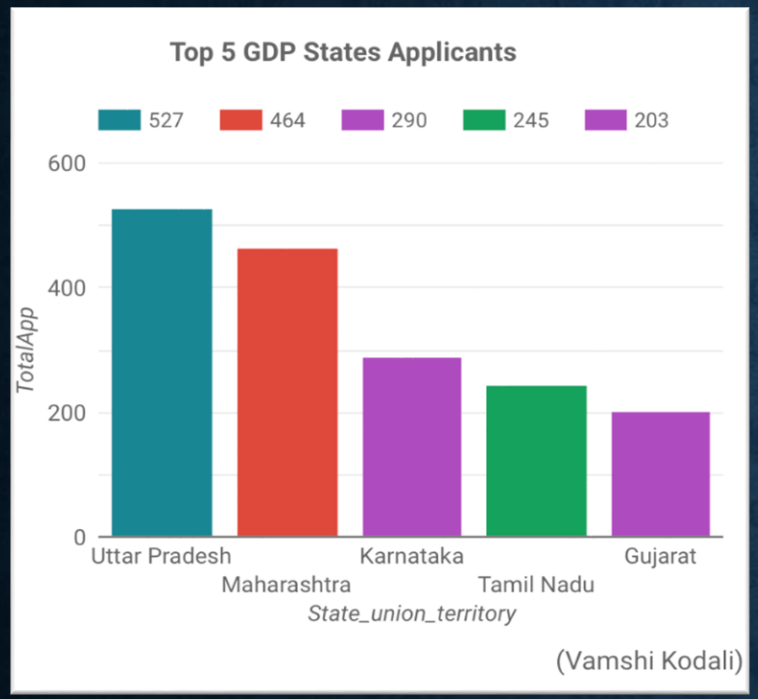
# ✓ CURRENT RESIDENCE GREATER THAN 2

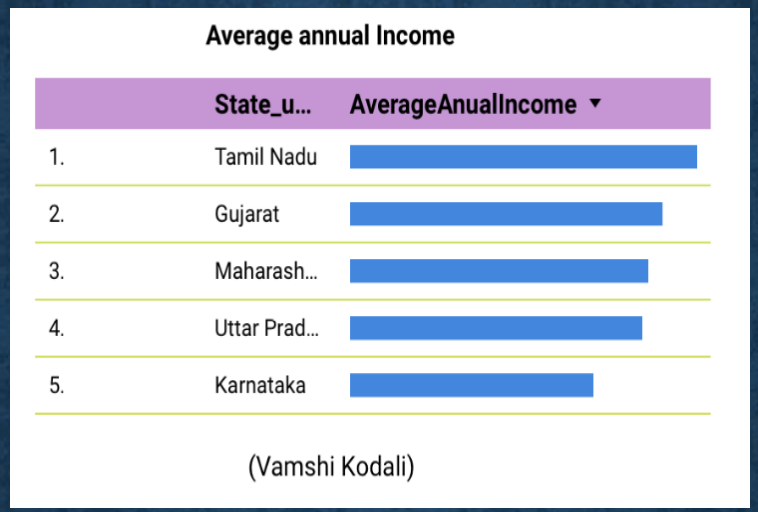
```
CREATE TABLE
saidineshkumarreddy_addula.Group8Project.CurrentResidenceGreater2 AS
SELECT
   State_union_territory,
   COUNT(*) AS TotalApp
FROM
   saidineshkumarreddy_addula.Group8Project.TopFiveStateApp
WHERE
   House_Ownership = 'owned'
   AND Years_in_Current_Residence >= 4
GROUP BY
   State_union_territory;
```

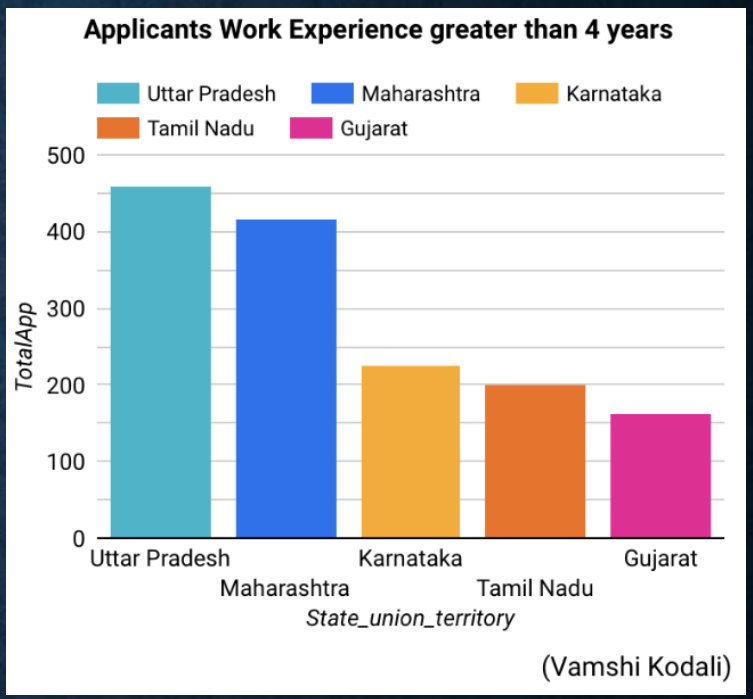



- The query generates a fresh table that displays the states or territories alongside the number of loan applications where the applicant possesses a home and has accumulated 4 or more years of work experience.

# ✓ VISUALIZATION



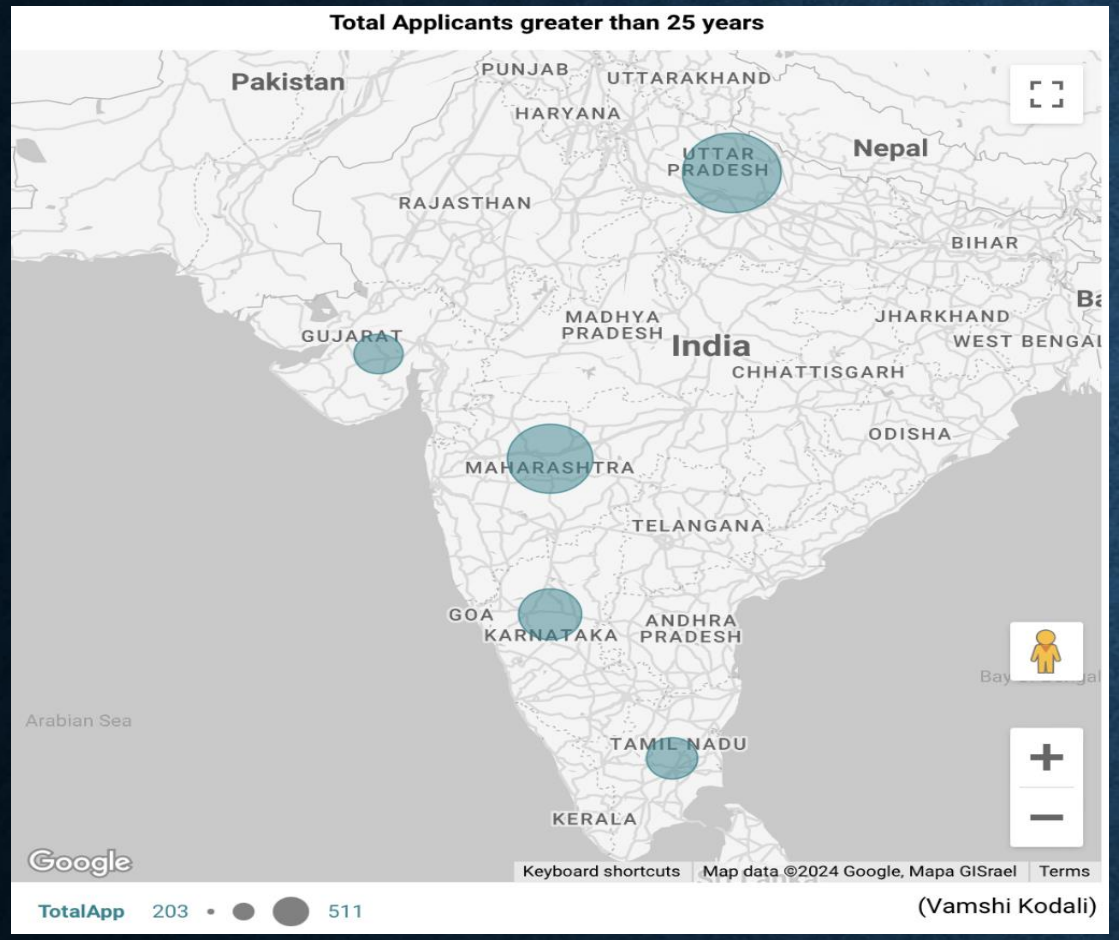




- The bar chart displays the number of applications from India's top 5 GDP states. Uttar Pradesh leads with 527 applications, followed by Gujarat (203), Tamil Nadu (245), Maharashtra (464), and Karnataka (464). The y-axis likely represents 'Total Applicants,' with state names on the x-axis.

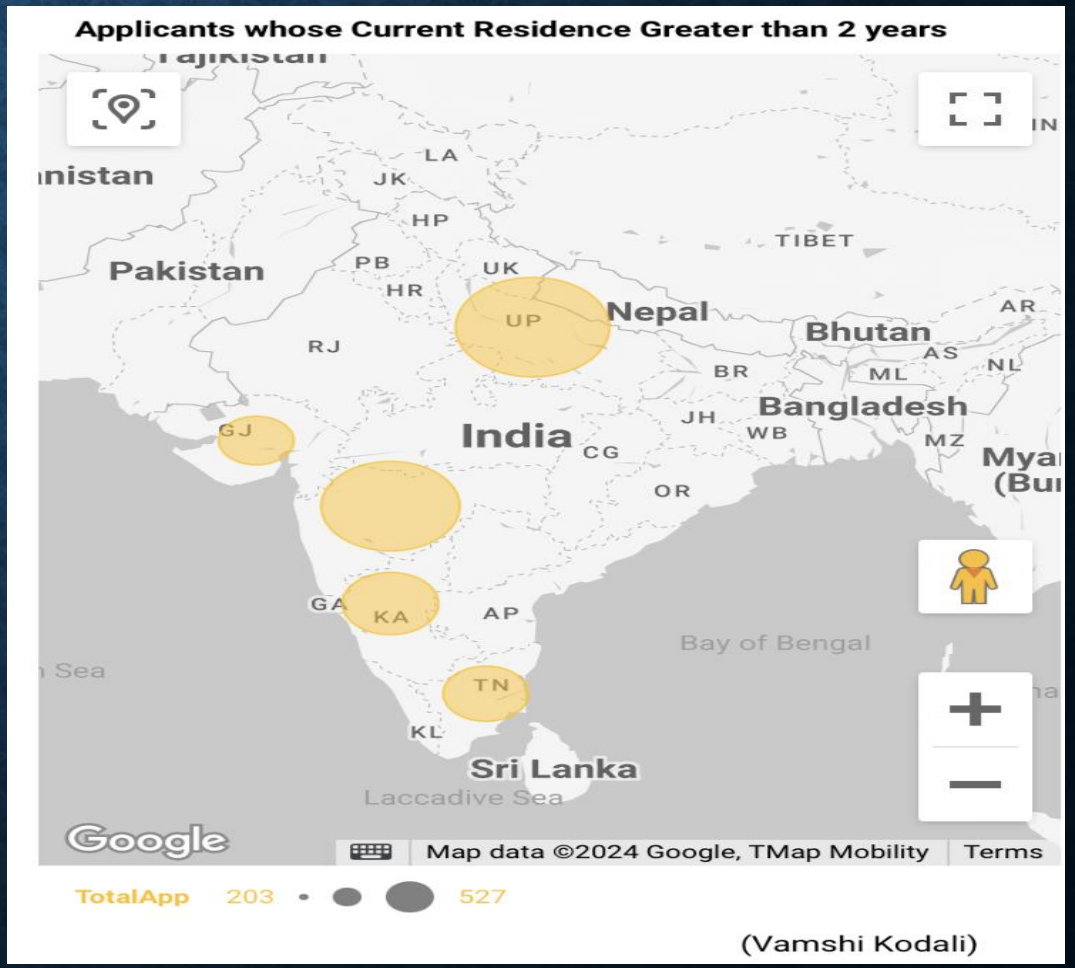
- The table displays the average annual income of the top five states in India: Tamil Nadu, Gujarat, Maharashtra, Uttar Pradesh, and Karnataka. Tamil Nadu has the highest income, while Karnataka has the lowest, with bars arranged in descending order accordingly.

- The chart illustrates applicants with over four years of experience from five Indian states. Uttar Pradesh leads with about 450, Maharashtra closely follows. Tamil Nadu and Karnataka have around 250-300 each, while Gujarat has about 150 applicants.




- The map of India shows circles over Gujarat, Maharashtra, Uttar Pradesh, Karnataka, and Tamil Nadu, representing the proportion of applicants over 25. The blue circles indicating relative numbers of applicants in each state.

- The map displays Indian applicants who've lived in the same place for over 2 years. Larger circles over Gujarat (GJ), Uttar Pradesh (UP), Karnataka (KA), and Tamil Nadu (TN) represent more applicants. After UP, Karnataka, Tamil Nadu, and Gujarat have the most applications.



## REFERENCE:


1. https://www.kaggle.com/datasets/yaminh/applicant-details-for-loan-approve
2. https://en.wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_GDP