ACPML - MySQL - Farmers Insurance Analysis - Executive Summary

By Deepak Kashyap, Kripa K, Sravan Muthe

Objective

This project aimed to assess the impact of India's Pradhan Mantri Fasal Bima Yojana (PMFBY) using structured agricultural insurance data. Using MySQL, the team performed comprehensive analysis across premiums, claims, coverage, and demographics to identify trends, regional gaps, and performance patterns.

Dataset Overview

The dataset spans multiple Indian states and years, organized into a normalized schema:

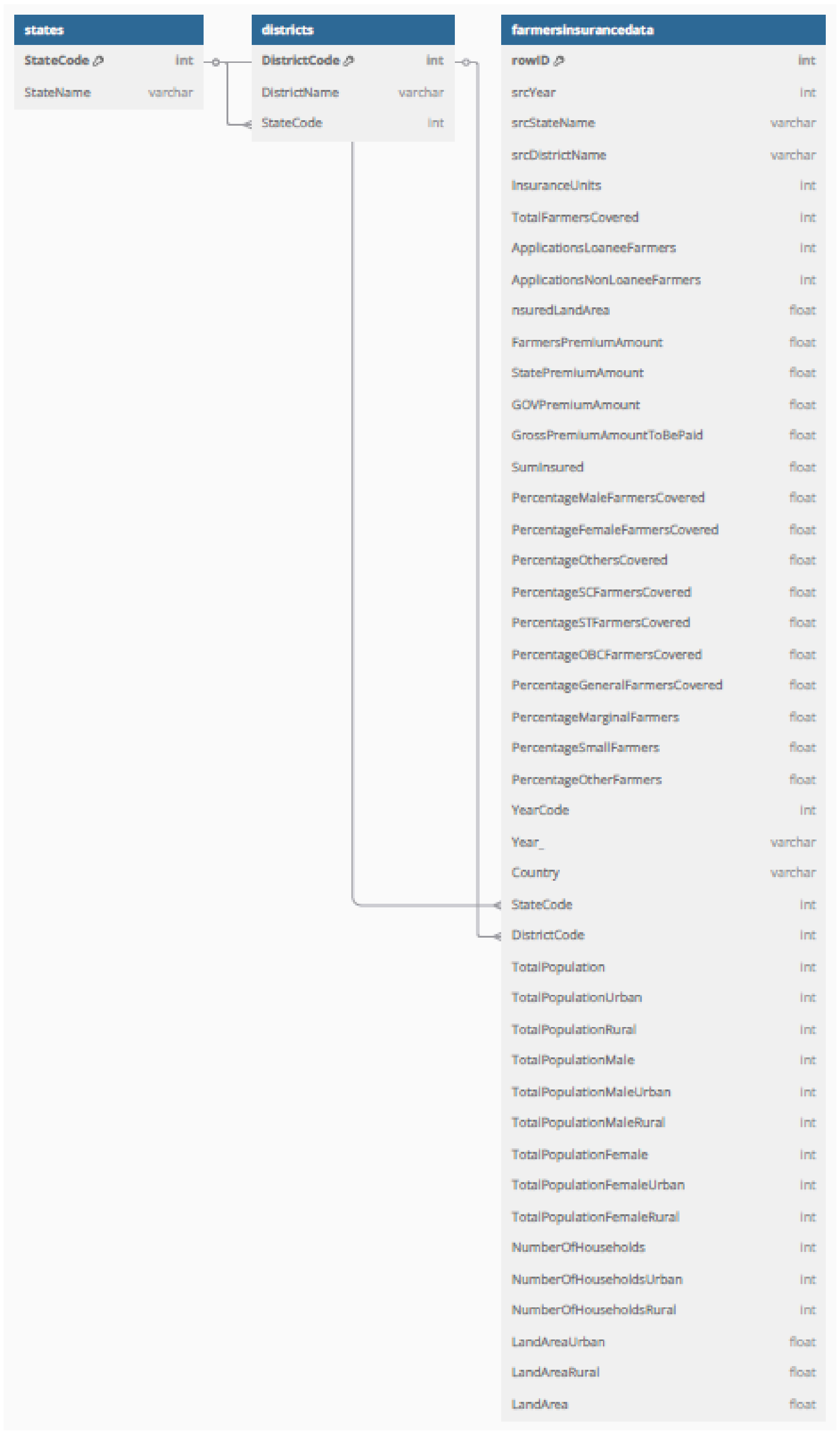
- - states — StateCode & StateName
- - districts — DistrictCode & DistrictName with FK to states
- - farmersinsurancedata — Core table with 40+ attributes including:

• Insurance metrics: SumInsured, FarmersPremiumAmount, InsuranceUnits

• Demographics: Population breakdowns, gender splits

• Location/time variables: srcStateName, srcDistrictName, srcYear

ER Diagram


states
StateCode int
StateName varchar
districts
DistrictCode int
DistrictName varchar
StateCode int
farmersinsurancedata
rowID int
srcYear int
srcStateName varchar
srcDistrictName varchar
InsuranceUnits int
TotalFarmersCovered int
ApplicationsLoaneeFarmers int
ApplicationsNonLoaneeFarmers int
nsuredLandArea float
FarmersPremiumAmount float
StatePremiumAmount float
GOVPremiumAmount float
GrossPremiumAmountToBePaid float
SumInsured float
PercentageMaleFarmersCovered float
PercentageFemaleFarmersCovered float
PercentageOthersCovered float
PercentageSCFarmersCovered float
PercentageSTFarmersCovered float
PercentageOBCFarmersCovered float
PercentageGeneralFarmersCovered float
PercentageMarginalFarmers float
PercentageSmallFarmers float
PercentageOtherFarmers float
YearCode int
Year_ varchar
Country varchar
StateCode int
DistrictCode int
TotalPopulation int
TotalPopulationUrban int
TotalPopulationRural int
TotalPopulationMale int
TotalPopulationMaleUrban int
TotalPopulationMaleRural int
TotalPopulationFemale int
TotalPopulationFemaleUrban int
TotalPopulationFemaleRural int
NumberOfHouseholds int
NumberOfHouseholdsUrban int
NumberOfHouseholdsRural int
LandAreaUrban float
LandAreaRural float
LandArea float

Schema Overview

Main Entity: AgriculturalInsuranceReport

Column Name	Data Type	Description
`rowID`	Integer	Unique row identifier
`srcYear`	Integer	Year of data collection
`srcStateName`	String	Name of the state
`srcDistrictName`	String	Name of the district
`Insurance units`	Integer	Units insured
`TotalFarmersCovered`	Integer	Total number of farmers covered
`ApplicationsLoaneeFarmers`	Integer	Applications from loanee farmers
`ApplicationsNonLoaneeFarmers`	Integer	Applications from non-loanee farmers
`InsuredLandArea`	Float	Total insured land area
`FarmersPremiumAmount`	Float	Premium paid by farmers
`StatePremiumAmount`	Float	State’s contribution to premium
`GOVPremiumAmount`	Float	Government contribution to premium
`GrossPremiumAmountToBePaid`	Float	Total premium
`SumInsured`	Float	Total sum invested

Demographic and Geographic Fields

Column Name	Data Type	Description
`TotalPopulation`, `Urban`, `Rural`	Integer	Population details
`TotalPopulationMale/Female`	Integer	Gender-specific population data
`NumberOfHouseholds`, `Urban`, `Rural`	Integer	Household data

Categorical and Percentage Fields

Column Name	Data Type	Description
`PercentageMaleFarmersCovered`	Float	% of male farmers covered
`PercentageFemaleFarmersCovered`	Float	% of female farmers covered
`PercentageSCFarmersCovered`, `ST`, `OBC`, `General`	Float	Caste/community-based percentages
`PercentageMarginalFarmers`, `Small`, `Other`	Float	Landholding category-wise percentages

Insights & Outcomes

Coverage & Participation

• Disparities in farmer coverage were evident across districts, with some rural zones still underinsured despite high population.

• Top-performing districts consistently covered >70% of their marginal and small farmers, while others lagged under 30%.

Premiums & Claims

• State and Government premium contributions varied widely, with some districts being heavily subsidized.

• Cumulative premium trends revealed seasonal spikes (e.g., kharif season) and underutilization in certain states.

Strategic Implications

• Findings can help redirect subsidies toward underinsured regions.

• Highlights potential for targeted awareness campaigns.

• Suggests need for improved claim outreach in districts with low farmer enrollment despite high disaster risk.

Final Analysis & Key Results

Sections 1–3: Basic Queries & Grouped Metrics

- States like Uttar Pradesh and Maharashtra had the highest overall count of insured farmers.

- In 2020, Tripura and Chhattisgarh stood out for their high farmer coverage ratio compared to the total population.

- Some districts in Himachal Pradesh had strong rural demographics but showed inconsistent premium participation.

- During 2018, districts such as Chikkamagaluru and Mysuru had zero reported premiums — indicating either low enrollment or system inefficiencies.

Sections 4–6: Sorted Trends, Text Functions, and Joins

- Thane, Pune, and Bangalore Urban had the highest total population among all districts.

- Certain districts with high insurance values recorded very low farmer premiums, implying strong public subsidies.

- Name pattern analysis showed several districts with suffixes like 'pur', highlighting regional naming clusters.

- Tripura consistently maintained strong coverage ratios over multiple years.

Sections 7–8: Subquery Insights & Analytical Functions

- Over 950 entries surpassed the national average for insured farmers.

- States such as Rajasthan had insured values far above districts with the highest premiums, showing possible risk balancing at the state level.

- Ranking logic (ROW_NUMBER, RANK) helped identify leading districts in both coverage and sum insured.

- The trend of cumulative premium growth was visible in states like Madhya Pradesh and Uttar Pradesh.

Sections 9–10: Schema Design & Data Edits

- Tables were designed with proper normalization: states, districts, and a central data table, all with relational integrity.

- Key data edits like UPDATE and DELETE were used effectively to simulate policy or data adjustments.

Conclusion

This study demonstrates the power of SQL in uncovering real-world patterns within public insurance schemes. It sheds light on underutilized regions, high-risk areas, and subsidy-driven zones, offering critical insights to inform agricultural policy and financial planning.

The *Farmers Insurance MySQL Case Study* served as a comprehensive platform to apply structured query language (SQL) to a real-world dataset representing India's agricultural insurance landscape. Leveraging a well-organized relational schema and strategically crafted queries, we uncovered valuable insights into state-wise and district-level insurance patterns, premium contributions, and demographic distributions.

The project strengthened our foundational SQL skills—such as filtering, aggregation, and joins—while also introducing advanced concepts including subqueries, window functions, and data normalization. Our analysis highlighted disparities in insurance coverage, identified high-risk or underinsured regions, and traced financial trends across different years and stakeholder groups.

As a team of three—Deepak, Kripa, and Sravan—we emphasized code quality, knowledge sharing, and efficient task division. The finalized ER diagram, normalized schema, and robust SQL query library exemplify both technical proficiency and practical relevance.