Final Analysis Report: Agricultural Insurance Dataset

Throughout this project, we conducted a comprehensive analysis of an agricultural insurance dataset that included information on farmers' coverage, premium payments, population statistics, and insurance metrics across multiple states and districts in India. The initial focus involved cleaning and standardizing the data-for example, updating the Year_ column for specific states like Himachal Pradesh and modifying individual premium records. A normalized schema was established by creating two interrelated tables: states, which included the state code and name, and districts, which incorporated district-level data along with a foreign key reference to the respective state. This helped enforce referential integrity and improve query efficiency.

We executed various types of SQL queries to extract meaningful insights from the dataset. These included filtering and deleting outdated or irrelevant records, such as removing entries with fewer than 10,000 farmers covered in 2020. Advanced SQL techniques like window functions (ROW_NUMBER(), RANK(), and SUM() over partitions) allowed us to rank districts, calculate cumulative premium trends, and identify the top-performing or underperforming regions over time. Using subqueries and joins, we compared premiums and insured values across regions, revealing interesting patterns-for instance, some districts had significantly higher premiums, while some states had larger insured sums than even the highest premium-paying districts.

One key insight was the noticeable variance in insurance adoption and premium payments across different districts and states. Some states showed consistently high farmers' coverage and sum insured values, indicating greater insurance awareness, risk management practices, or agricultural dependency. We also observed districts where farmers paid significantly higher premiums than the state average, which could point to localized risk factors or intensive cropping practices. The combination of insurance data with demographic statistics-such as total population and number of households-allowed for a richer understanding of farmer participation trends.

In conclusion, the dataset has been cleaned, structured, and analyzed to reveal both macro and micro-level insights. The resulting SQL scripts support accurate reporting, dashboard creation, and are ready to be extended for advanced analytics, including predictive modeling. This foundation sets the stage for policy optimization, regional planning, and future data science initiatives centered on agricultural insurance and rural development.