Date : 11-11-2025

Data Preprocessing

Project Name	Total revenue	No of emp	Per person revenue
New Satelite launch	136400	6	22733
Janm Hit	120800	4	30200
Invest India	299100	2	149550
Expand Reach	184400	3	61466
Self Employ	246600	3	82200
Connect large	281100	4	70275
5G All	74800	2	37400
Tech Launc	214000	4	53500
Automate Sales	99200	5	19840
Install Tower	293100	3	97700
Make in India	127200	3	42400

Project Name	Sectors	Total Project Revenue	Name	Job role
New Satelite launch	IT	136400	karan ram	analyst
New Satelite launch	IT	136400	vijay roy	analyst
New Satelite launch	IT	136400	arjun joshi	software
New Satelite launch	IT	136400	krishna das	mis
New Satelite launch	IT	136400	laxmi nayak	analyst
New Satelite launch	IT	136400	ganesh	support
Janm Hit	Service	120800	darshana nair	mis
Janm Hit	Service	120800	rajkumar joshi	analyst
Janm Hit	Service	120800	ajay kumar	mis
Janm Hit	Service	120800	ganesh	support
Invest India	Service	299100	disha jhah	analyst
Invest India	Service	299100	rohini patak	software
Expand Reach	Telecom	184400	jay naik	analyst
Expand Reach	Telecom	184400	krishna das	mis
Expand Reach	Telecom	184400	ajay kumar	mis

Insights

From the cleaned and grouped project data, it is observed that projects like **Invest India**, **Connect Large**, and **Self Employ** generate comparatively higher revenue per employee, indicating better efficiency and higher-value contribution roles in those teams. Projects with **larger teams but lower per-person revenue** may involve support or shared operational tasks, showing moderate productivity levels. The **IT and Telecom sectors** hold the highest number of projects, suggesting organizational focus on technology-driven and communication-based initiatives. The shortlisted team summary also highlights varied job roles such as **Analyst, MIS, Software, and Support**, showing that each project's success depends on a balanced team skill structure.

Steps

1. **Imported the raw Excel dataset** into the workspace for cleaning and preparation.
2. **Identified the combined text columns** where multiple values (name, role, etc.) were grouped together.

3. **Applied 'Split Column by Delimiter (Comma)'** to separate combined values into individual structured fields.
4. **Renamed the resulting columns** to meaningful and standardized names such as: *Project Name, Sector, Total Project Revenue, Employee Name, Job Role.*
5. **Removed unnecessary and empty rows** to ensure only relevant records remained.
6. **Checked for and removed duplicate entries** to maintain data accuracy and avoid double counting.
7. **Standardized text values** by converting all fields to lowercase and trimming extra spaces to avoid mismatches.
8. **Replaced inconsistent naming values** (e.g., variations in job roles like *Mis, MIS, mis* standardized to *MIS*).
9. **Duplicated the cleaned table** to perform calculations without altering the original cleaned dataset.
10. **Grouped the duplicated table by Project Name** to summarize revenue and employee counts per project.
11. **Calculated Total Revenue** for each project by summing up the revenue column.
12. **Calculated Number of Employees per Project** to understand team size distribution.
13. **Derived Per Person Revenue** using the formula: *Per Person Revenue = Total Revenue ÷ Number of Employees.*
14. **Created a Project Team Summary** that lists employees and their job roles assigned to each project.
15. **Verified consistency** across all datasets to ensure no mismatch in project names or sectors.
16. **Final datasets prepared** are now ready for visualization, sector-wise comparison, and performance insights.