Oil Production Data Analysis Report

1. Dataset Description

1.1 Source: Oil production dataset (covering multiple companies and refineries, spanning years 2020–2024).

1.2 Columns:
- Month – Month of production record
- Year – Year of production record
- OIL COMPANIES – Company/refinery name
- Quantity – Oil production quantity (in 000 Metric Tonnes)

1.3 Data Quality:
- No major missing/null values reported
- Dataset schema is consistent
- Covers both company-level and aggregate production

2. Operations Performed

2.1 Data Cleaning & Exploration
- SparkSession created and dataset loaded successfully
- Verified schema and data types
- Converted dataframes for visualization

2.2 Descriptive Analytics
- Total production by oil company computed
- Year-wise total production aggregated
- Monthly variations analyzed
- Visualizations created: bar charts, violin plots, company comparisons

2.3 Relationship Analysis
- Compared production across top oil companies
- Observed trends over time (2020–2024)
- Analyzed seasonal/monthly patterns

3. Key Insights

3.1 Production Trends
- Total production shows strong growth from 2020 (468,578) to 2023 (761,508)
- 2024 shows partial data (194,552 so far)
- Indicates consistent increase in demand and supply capacity

3.2 Company-Level Insights
• RIL (Reliance Industries Ltd.) is the largest contributor (~249,000+ MT)
• IOCL and BPCL follow as major players
• Certain refineries (e.g., RIL Jamnagar, BPCL Kochi, IOCL) dominate production

3.3 Geographic/Refinery Insights
• Jamnagar refineries (RIL) lead production scale
• Kochi (BPCL) and Gujarat sites contribute significantly
• ONGC and HPCL production is moderate

3.4 Temporal Insights
• Clear annual growth trend (2020 → 2023)
• Monthly variations suggest seasonality or operational cycles

4. Recommendations

4.1 Production Planning
• Forecast demand for 2025 based on annual growth trend
• Ensure capacity expansion to meet rising demand

4.2 Efficiency Improvement
• Smaller companies should optimize operations to compete with major refineries
• Benchmark processes against RIL/IOCL leaders

4.3 Risk Diversification
• Avoid over-dependence on few refineries (Jamnagar, IOCL)
• Strengthen regional distribution of production

4.4 Future Analytics Opportunities
• Build predictive models for production forecasting
• Analyze correlations with demand, prices, and seasonal trends
• Create dashboards for real-time monitoring and decision support