Project Overview

This project demonstrates an automated, LLM-driven workflow for generating structured analytical reports from structured datasets, using synthetic BMW sales data as an example. The system leverages a **LangGraph workflow** to orchestrate multiple autonomous agents powered by the Gemini 2.5 Flash LLM. Each agent executes a specific analytical task such as data preview, visualization, key driver analysis, or performance ranking. LangGraph coordinates the execution order, manages data flow, and compiles the final report into a **Word document (DOCX)**.

The goal is to illustrate a prototype where a language model functions as a central orchestrator, reasoning and acting across multiple steps to produce structured, actionable reports. While the case study focuses on sales data, the architecture and tools are generalizable to other structured datasets and business reporting tasks.

System Architecture

The system consists of three main components

1. **LLM Orchestrator with LangGraph**

   Gemini 2.5 Flash serves as the core LLM. It has a lightweight design and robust support for tool calling. Using LangGraph, the workflow is structured as a graph of nodes. Each node represents a distinct analytical agent.

   The **LangGraph workflow structure** is as follows:

   - **Four parallel agents** perform analysis simultaneously:
     1. Sales Trend Agent
     2. Top and Under Performer Agent
     3. Key Driver Agent
     4. Generic Analysis Agent
   - Outputs from all four agents feed into a **Report Writer Agent**
   - The Report Writer Agent compiles all outputs and generates a final **Word document (DOCX)**
   - The workflow ends after the DOCX report is produced

2. This structure allows the LLM to execute multiple tasks in parallel, iterate or refine results, and produce a high-quality report automatically.

3. **Reusable Python Tool Suite**
   The system includes tools for data preview, visualization, performance ranking, OLS regression for key driver analysis, and DOCX report assembly. Tools are **dataset-agnostic**, enabling adaptation to new domains. Each tool provides structured input and output formats for autonomous LLM use.

4. **LangGraph Nodes and Agents**

    - **Sales Trend Agent**: Summarizes overall sales trends and generates line and bar plots

    - **Top and Under Performer Agent**: Identifies high- and low-performing products, regions, or models

    - **Key Driver Agent**: Performs OLS regression to quantify relationships between features and outcomes

    - **Generic Analysis Agent**: Explores additional hypotheses and performs combined analyses

    - **Report Writer Agent**: Compiles agent outputs into a structured Word document, embedding plots and narrative text

---

Tooling and Functionality

- **Automated Visualization**: Agents generate line and bar charts with group-level breakdowns. Plots are linked to underlying data, enabling the LLM to produce narrative summaries consistent with quantitative evidence

- **Statistical Analysis**: The OLS regression tool identifies key drivers, handles categorical variables through one-hot encoding, and outputs coefficients, p-values, and R-squared metrics

- **Automated Report Assembly**: The Report Writer Agent takes structured JSON input, embeds plots, and produces a polished DOCX report with coherent narrative sections

---

LLM Orchestration and Autonomy

- LangGraph nodes enable the LLM to execute multiple agents in parallel or sequence
- Outputs from one agent feed into others, supporting iterative refinement and error correction
- The workflow mimics the decision-making process of a human data scientist, integrating reasoning, acting, and report synthesis into a unified pipeline

---

Extensibility and Creative Analysis

- Beyond pre-defined agents, the workflow supports exploration of new hypotheses, additional statistical analyses, and generation of novel recommendations
- The modular design allows the workflow to adapt to new business questions or datasets without code changes

---

Business Insights and Impact

- **Performance Trends**: Identification of top and under-performing models, regions, and markets
- **Sales Drivers**: Quantified relationships between features and outcomes via regression analysis
- **Actionable Recommendations**: Strategic guidance for pricing, product focus, and marketing decisions

By automating reporting, this workflow reduces manual effort, accelerates decision-making, and provides a reproducible, standardized framework for data-driven analysis

---

Conclusion

This project demonstrates how LangGraph combined with an LLM and modular tools can fully automate complex business reporting. The architecture enables parallel agent execution, iterative reasoning, visualization, and **direct DOCX report generation**.