

INTERNSHIP REPORT: Advanced Sales Analytics using - Google Looker Studio

Name: Aditya Bhagwat Basude

Internship Position/Title: Data Analyst Intern

Company: Null Class

Date of Submission: 10-12-2025

1. Introduction

During my internship, I worked as a Data Analyst, handling large datasets and performing end-to-end data analysis using Python, Pandas, NumPy, Matplotlib, Seaborn, and MySQL. The primary objective was to solve 19 business-driven analytical tasks related to sales, marketing, customer behavior, forecasting, and operational insights.

This report summarizes my work, learnings, challenges, solutions, and overall impact during the internship period.

2. Background

The organization provided multiple raw datasets, these datasets were stored in MySQL and exported as CSV files. My role was to merge, clean, preprocess, analyze, visualize, and derive insights that supported various internal teams like Marketing, Sales, Finance, Operations, and Warehousing.

The tasks included customer segmentation, time-series forecasting, discount impact analysis, profit evaluation, category-level performance, and many more.

3. Learning Objectives

The internship aimed to help me develop skills in:

- Data extraction from SQL databases
- Data cleaning and preprocessing
- Merging large relational datasets
- Exploratory Data Analysis (EDA)
- Business-driven analytics
- Statistical calculations and KPIs
- Predictive modeling (Regression forecasting)
- Data visualization using Matplotlib & Seaborn
- Professional reporting and insight generation

4. Activities and Tasks

The core of the internship involved performing end-to-end sales analytics across 19 tasks, covering data preparation, business problem solving, and visualization.

- **Data Merging & Cleaning:** Combined multiple CSV files exported from MySQL, resolved column conflicts (e.g., $id \rightarrow sku_id$, $id \rightarrow customer_id$), standardized date formats using `pd.to_datetime()`, and created additional fields such as `year`, `month`, `day_name`, `month_name`, `net_profit`, and `discount_category` to support deeper analysis.
- **Product Performance Analysis:** Identified the top products in categories such as *Mobiles & Tablets (2022)* and evaluated declining product performance—especially focusing on the *others* category across 2021–2022.
- **Campaign & Trend Evaluation:** Analyzed weekend vs weekday performance during *Q4 2022* promotional campaigns and studied month-wise sales trends using aggregated time-series insights.
- **Customer Behaviour Insights:** Calculated the time taken for customers to convert from registration to first order, segmented customers by profitability, and assessed repeat purchase patterns.
- **Revenue & Discount Impact:** Compared revenue before and after discounts across product categories, measured discount effectiveness, and evaluated payment method performance.
- **Forecasting & Predictive Analysis:** Built a regression-based model to forecast sales for *Q2 2023* using historical monthly data, computing MAE and RMSE for accuracy validation.
- **Visualization Development:** Created a variety of analytical visuals including time-series line charts, clustered and horizontal bar charts, histograms, pie charts with external labels, and dual-axis combo charts to communicate insights effectively to marketing, sales, finance, and operations teams.

A significant aspect of the workflow involved using **head ()** and **head (12)** of similar during development to validate logic on small subsets before running computations on the full dataset, ensuring accuracy and efficiency throughout the analysis process.

5. Skills and Competencies

Technical Skills

- **Python:** Pandas, NumPy, Matplotlib, Seaborn
- **SQL:** Querying, joins, grouping, filtering
- **Data Cleaning:** Handling missing values, duplicates, data type conversions
- **Data Modeling:** Linear regression forecasting
- **Data Visualization:** Professional dashboard-style charts
- **Statistical Analysis:** Growth %, MAE, RMSE, average time-to-conversion

Soft Skills

- Analytical problem-solving
- Business communication
- Attention to detail
- Report writing and interpretation
- Time management

6. Feedback and Evidence

During the tasks, I continuously validated my work by checking:

- DataFrame structures using .info()
- Sample data using .head() and .head(12)
- Unique values using .unique()
- Summary statistics using .describe()

Why I used head () and head (12) frequently

Because the datasets were **very large**, running visual inspections on the entire dataset would:
Slow down the notebook, cause unnecessary computation, Make debugging harder

Therefore, I used:

- df.head() → to quickly check column correctness

- `df.head(12)` → to inspect a full month cycle or ensure grouping logic is correct

This approach helped me confirm that the **code was working properly before applying it to the entire dataset**, reducing the risk of errors.

This was a key part of my workflow and improved efficiency significantly.

7. Challenges and Solutions

Challenge 1: Large datasets caused slow execution

- **Solution:** Used `head()`, `sample()`, filtered subsets, and incremental validation.

Challenge 2: Column name conflicts after merging

- Columns such as `id` appeared in multiple tables.
- **Solution:** Renamed keys before merging (`id` → `sku_id`, etc.).

Challenge 3: Datatype mismatches when merging

- Example: `payment_id` stored as string in one file and integer in another.
- **Solution:** Forced consistent types using `.astype(str)`.

Challenge 4: Missing or invalid dates

- Some records had null values.
- **Solution:** Used `errors="coerce"` in `pd.to_datetime()` and removed invalid rows.

Challenge 5: Forecasting errors due to volatile data

- Sales had spikes and drops causing high MAE/RMSE.
- **Solution:** Still used linear regression as required, and provided interpretation about volatility.

8. Outcomes and Impact

The analyses delivered had a direct impact across multiple teams. Marketing gained insights into top products, discount performance, campaign effectiveness, and customer conversion patterns. Sales benefited from quarterly sales forecasts, category trends, and identification of declining products. Operations and Warehouse teams used the findings to understand inventory risks and slow-moving items, while Finance evaluated payment method performance, revenue impact of discounts, and customer profitability.

Overall, the insights supported better decision-making across the organization, improving budget planning, inventory management, campaign optimization, customer retention strategies, and revenue growth efforts.

9. Conclusion

This internship provided a strong foundation in **real-world data analytics**, allowing me to work with complex datasets, business problems, SQL queries, predictive modeling, and visualization techniques. I developed both technical and analytical skills and gained hands-on experience in transforming raw data into actionable insights.

The tasks completed improved my confidence in working independently as a data analyst, and the experience significantly enhanced my ability to handle end-to-end analytical workflows.

I express gratitude to the organization for providing this valuable learning experience.