final_project_part_one

August 27, 2024

0.1 FINAL PROJECT: The Plan

Cesar Chaparro - DA Student

0.1.1 Client: AtliQ Hardware.

0.1.2 AtliQ Hardware is one of the leading computer harware producers in India, and has even expanded into more than 7 countries as well. They sell different types of hardware to big players such as Amazon, Best Buy, and Walmart.

0.1.3 This year, they're asking PWC to conduct a big audit of their sales and help them automate their existing data.

0.1.4 FINNANCIAL ANALYSIS

DECOMPOSITION AND RESEARCH TASKS:

0.1.5 1. Definition of Objectives and Key Questions

Objective: Understand how revenue, profits, and margins have changed over time. Analyze shifts in the market and identify the most profitable categories.

Key Questions:

- How have revenue and profits trended over the years?
- What are the profit margins, and how have they evolved?
- Which markets, platforms, or channels have driven the most revenue and profit?
- What categories of products are the most profitable?
- How have different regions/subzones contributed to revenue and profit?
- Are there any noticeable shifts in market trends or customer preferences?

0.1.6 2. Identify the Important Metrics

Revenue Metrics:

- Total Revenue: Sum of gross_price * sold_quantity from fact_sales_monthly.

- Revenue Growth Rate: Percentage increase/decrease in revenue year-over-year. #### Profit Metrics:
- Gross Profit: Revenue minus the manufacturing cost (gross_price * sold_quantity - manufacturing_cost * sold_quantity).
- Profit Margin: Gross profit as a percentage of total revenue.
- Net Profit: Gross profit minus pre-invoice discounts (gross_profit - (pre_invoice_discount_pct * gross_profit)).
- Net Profit Margin: Net profit as a percentage of total revenue. #### Market and Category Metrics:
- Revenue by Market: Revenue broken down by market.
- Revenue by Platform: Revenue broken down by platform.
- Profit by Category: Profitability of different segment and category. #### Trend Metrics:
- Market Share: Proportion of total revenue by market.
- Revenue/Profit by Year: Trend of revenue and profit over time (fiscal_year).

0.1.7 3. Data Preparation and Cleaning

Data Integrity Check:

- Ensure each market belongs to one sub_zone and region.
- Validate that each customer has only one associated platform.
- Check that product_code is consistent across all fact tables.
- Verify that there are only 3 unique values in the division column of the dim_product table.
- Verify that there are only 6 unique values in the segment column of the dim_product table.
- Verify that each single market belongs to a one subzone and one region only. #### Handle Missing Data:
- Identify and address missing values in key columns like gross_price, manufacturing_cost, and sold_quantity.
- Impute or remove records with missing data as appropriate. #### Feature Engineering:
- Create new features like gross_profit, net_profit, profit_margin, and net_profit_margin for further analysis.
- Aggregate data by fiscal_year, market, platform, etc., to facilitate trend analysis.

0.1.8 4. Exploratory Data Analysis (EDA)

Descriptive Statistics:

- Calculate mean, median, and standard deviation for revenue, profit, and margin metrics.
- Identify outliers or unusual trends in the data. #### Trend Analysis:
- Plot revenue, profit, and margin over time to visualize trends.
- Use time series analysis to forecast future revenue and profit. #### Segment Analysis:
- Break down revenue and profit by platform, market, region, and category.
- Identify which segments contribute most to overall profitability. #### Correlation Analysis:
- Analyze the relationship between different metrics (e.g., how does discount percentage affect net profit?).
- Look for correlations between market conditions and profitability.

0.1.9 5. Identifying Problems or Questions to Address

Profitability Issues:

- Are certain markets or platforms less profitable? If so, why?
- Is there a declining trend in any key categories or segments? #### Market Shifts:
- Are there shifts in revenue from one platform to another (e.g., from Brick & Mortar to E-Commerce)?
- Are there emerging markets or declining markets based on the data? #### Cost Efficiency:
- Are manufacturing costs rising, and how does this affect profit margins?
- Are discounts eroding profitability in certain markets or for certain customers?

0.1.10 6. Research and Contextual Understanding

Industry Benchmarks:

- Research industry standards for profit margins, cost structures, and market share to compare with your findings. #### Economic Factors:
- Consider external economic factors that may have influenced market trends (e.g., currency fluctuations, economic downturns). #### Competitor Analysis:
- Research competitors' performance in similar markets to understand potential threats or opportunities.

0.1.11 7. Hypothesis Formation and Testing

Based on the EDA, form hypotheses about what drives revenue and profit. For example:

- "E-Commerce platform is more profitable due to lower overhead costs."
- "The market in Japan has seen a decline due to increased competition." #### Design experiments or statistical tests to validate these hypotheses.

0.1.12 8. Visualization and Reporting

Dashboards:

- Create visualizations to represent the trends, comparisons, and insights derived from the data.

Report:

- Compile a report summarizing findings, insights, and recommendations based on the analysis.
- Ensure that the report answers the key questions posed at the beginning of the project.

0.1.13 9. Actionable Recommendations

Based on the analysis, provide recommendations for business strategies to improve revenue, optimize profits, and adapt to market shifts. These could include:

- Focusing more on profitable platforms or markets.
- Reducing costs in certain product lines.
- Increasing marketing efforts in emerging markets or declining segments.

0.1.14 10. Tools used through the project

- SQL (SQLite): Use SQL databases to store, query, and manipulate large datasets efficiently.

- Pandas (Python): For in-memory data manipulation and analysis. Pandas for loading data from SQL databases, performing data cleaning, transformation, and aggregation.

- Matplotlib & Seaborn (Python): For generating plots and visualizations. These libraries are useful for visualizing trends, distributions, and relationships in the data.

- Tableau: For creating interactive dashboards to visualize trends and insights from the data. They are excellent for communicating findings to stakeholders.

- Markdown (Jupyter Notebooks): To document my process and findings within Jupyter Notebooks using Markdown cells.

- Git/GitHub: For version control, allowing to track changes to all the code, data queries, and reports. Specially useful to work in a team.
