High Level Design

BUDGET SALES ANALYSIS

MANISHA RAJ A.

DOCUMENT VERSION CONTROL

Data Issue	Version	Description	Author
	1.0	First version of Complete HLD	Manisha Raj A.

Table of Contents

ABSTRACT

The resale business is growing rapidly in densely populated cities, leading to increased competition in the market. This document uses a dataset from Adventure Works, which includes three years of historical sales data. Implementing effective data analysis systems can enhance the company's performance and deliver better returns on investment for stakeholders.

I. INTRODUCTION

I.I PURPOSE OF HIGH-LEVEL DESIGN DOCUMENT

This High-Level Design (HLD) Document aims to provide detailed information about the project, serving as a guide for coding. It helps identify any inconsistencies before development begins and can act as a reference for understanding how different parts of the system will interact.

The HLD will:

- Outline all design elements in detail
- Describe the user interface
- Explain hardware and software interfaces
- Specify performance requirements
- Detail the project's architecture and design features
- List and describe important non-functional aspects, including:
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

I.2 SCOPE.

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2. General Description.

2.1 Product Perspective and Problem Statement.

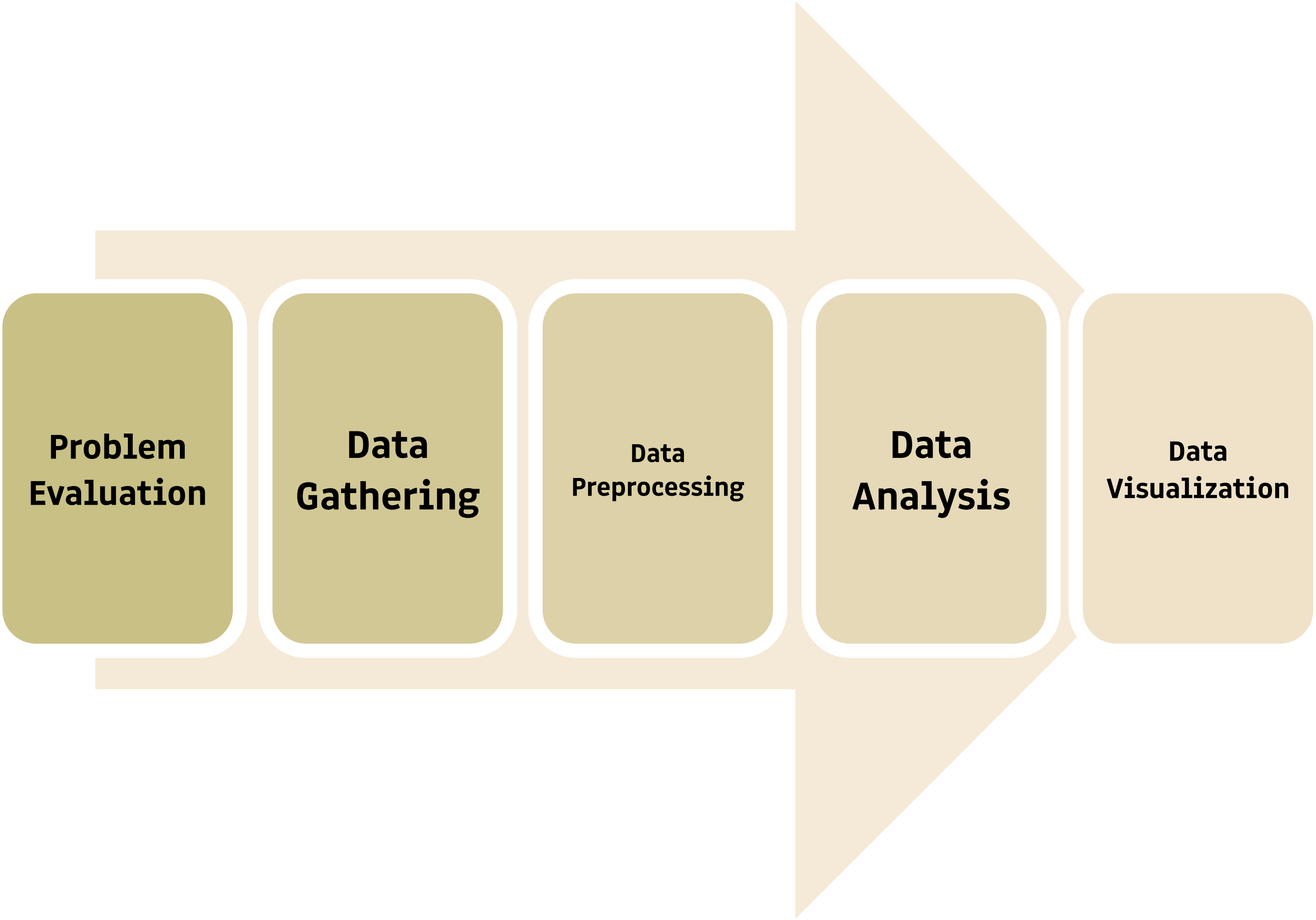
The aim of this project is to analyze sales data and assess the sales team's performance against targets. It will identify both high-performing and low-performing products and services, highlight challenges in meeting goals, and reveal market opportunities and sales activities that drive revenue.

2.2 TOOLS USED

The framework is built using various Business Intelligence tools and libraries, including NumPy, Pandas, Seaborn, Matplotlib, MS Excel, MS Power BI, Jupyter Notebook, and the Python programming language.

3. DESIGN DETAILS.

3.I FUNCTIONAL ARCHITECTURE.



3.2 HOW BI WORKS

Organizational Memory	Information Integration	Insight Creation	Presentation
Data Warehouse	Business Analytics Tool	Text Mining tools	OLAP Tools
Enterprise Resource Planning (ERP)	Data Mining	Web Mining tools	Visualization tool
Knowledge Repository	Real-time Decision	Environmental scanning	Digital Dashboard
Content Management System (CMS)		RFID	Score cards

3.3 OPTIMIZATION.

1.Data Loading Optimization:

- Efficient Data Loading: The project utilized direct links to Excel files stored in GitHub repositories, enabling dynamic and real-time data updates. This removed the need for manual uploads, reducing the chance of outdated data.
- Selective Sheet Loading: Instead of loading entire Excel files into memory, only the required sheets (such as Calendar, Sales, Customers, Products, and Territory) were imported. This selective data loading minimized unnecessary memory usage.

5.2 Data Cleaning Optimization:

- Batch Cleaning for Budget Data: The first two metadata rows of the Budget dataset were removed, and column headers were adjusted to ensure proper alignment with data fields. This reduced overhead by preventing additional cleaning operations later in the process.
- Dropped Redundant Columns: Columns that added little analytical value (e.g., Unnamed and cost-related fields) were dropped at the initial stages, saving both memory and processing time during further operations. This step streamlined subsequent data merging and processing.
- Handling Missing Values: A threshold of 50% missing values was applied, leading to the efficient removal of columns that would have negatively impacted data integrity and analytical quality.

3. Data Merging Optimization:

- Efficient Merging Strategy: DataFrames were merged using pd.merge on key columns (such as CustomerKey and ProductKey) in a stepwise manner. This reduced unnecessary duplication and ensured optimal memory usage, avoiding issues with redundant data merges.

4. Exploratory Data Analysis (EDA) Optimization:

- Smart Data Visualization: Only key insights such as gender distribution, occupation vs. sales amount, and commute distance vs. sales amount were visualized using Seaborn and Matplotlib. This ensured that the visualizations focused on providing actionable insights rather than overwhelming the dashboard with redundant information.
- Use of Color Palettes: Optimized color palettes such as 'coolwarm,' 'magma,' and 'inferno' were chosen for visualizations to ensure that patterns and trends were more easily identifiable by stakeholders.

5. Performance Tuning:

- Efficient Memory Usage: Dropping irrelevant columns and rows at the data cleaning stage significantly reduced memory footprint. By eliminating unnecessary fields early on, the processing of data became faster, enabling smoother execution of downstream tasks like EDA and reporting.
- Optimized Computation of Aggregates: Grouping operations were efficiently performed using groupby functions, reducing the time complexity of aggregate calculations (such as sales totals by occupation, education level, and product category).

4. KPI.

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the sales As and when the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

Key indicators displaying a summary of the sales generation and its relationship with different metrics

Growth Analysis:

- Quarterly Growth/Decay Rate
- Seasonal Growth/Decay Rate
- Monthly Growth/Decay

Sales Analysis:

- Year-Wise Sales Comparison
- Monthly Sales and Profit
- Monthly Transactions

Sales Performance Report:

- Total Sales
- Country-Wise Sales
- Top 10 Customers by Sales
- Profit Margin
- Average Daily Sales

Country-Wise Analysis Dashboard:

- Total Sales by Category
- Sub-category Sales

Category-Wise Analysis Dashboard:

- Total Sales
- Average Daily Sales
- Growth/Decay Rate
- Profit Margin
- Total Sales by Country

Customer Profile Dashboard:

- Customer
- Customer Key
- Age
- Annual Income
- Category-Wise Sales:
- Bikes
- Clothing
- Accessories

5. Deployment

- Prioritizing data and analytics is more important than ever. Regardless of your company's size, you're already gathering data and likely only analyzing a small portion of it to address business issues, gain a competitive edge, and drive overall transformation.

- With the rapid increase in enterprise data and database technologies, and the high demand for analytical skills, today's most successful IT organizations are focusing on enabling self-service. They do this by using Power BI on a large scale and by organizing and integrating various data sources. This allows both business users and experts to easily create and access useful content.

- Power BI Desktop and Power BI Service make the most of your existing technology investments and integrate them into your IT infrastructure. They offer a modern, self-service analytics platform that can be tailored to your needs, with options available for on-premises, cloud, and hosted environments.