Executive Summary: Financial Clustering & Anomaly Detection Dashboard

1. The Problem: Uncovering Meaningful Patterns in Complex Financial Data

Financial analysts and decision-makers often struggle with identifying consistent patterns, risk signals, or profitable segments within large, noisy datasets. Publicly available financial datasets contain valuable, yet underutilized, insights about company performance, market trends, and potential anomalies.

Our group project aimed to apply the **full data pipeline** — from **ETL (Extract, Transform, Load)** to **Data Mining** and ultimately to **Insights and Storytelling**. We cleaned a dataset comprising time-series and financial metrics across multiple companies, with the goal of uncovering trends, clusters of similar behavior, and outliers that may represent risk or opportunity.

By combining **clustering techniques** and **anomaly detection**, we sought to bring clarity to the complex, high-volume financial data, enabling stakeholders to make informed, data-backed decisions.

---

2. Recommended Solution: Data Mining with Visual Insight Delivery

Our solution was a multi-step data mining process, summarized as follows:

- **KMeans Clustering** was used to group companies based on key financial indicators such as profit margin, volatility, and stock behavior.
- **Time-based Anomaly Detection** was applied to identify companies showing unusual patterns or potential red flags.
- **A Power BI Dashboard** was built to allow users to explore clusters, anomalies, company trends, and distributions interactively.
- **Visual storytelling** enabled us to highlight 3–5 key insights derived from the mining process.

---

3. Solution Value: Clear Insights, Targeted Recommendations

Our solution delivered practical business value by translating raw data into **actionable insights**. Key findings include:

- **Cluster 2 companies** demonstrated the highest average profit margins and consistent performance, signaling potential low-risk, high-reward investment opportunities.

```
Cluster
0    14.601414
2    13.889606
1     4.742146
Name: Profit_Margin,
```

- **Anomalies were concentrated around certain time periods**, suggesting either market-wide volatility or company-specific events needing deeper investigation.

```
Number of anomalies detected: 10
          Date   Company  Stock_Price  Profit_Margin
0   2022-01-01  EnergyCo         65.3      12.867542
119 2022-01-01  RetailCo         42.5       4.373798
120 2022-02-01  RetailCo         43.1       4.413027
121 2022-03-01  RetailCo         42.8       4.385095
122 2022-04-01  RetailCo         43.5       4.435534
157 2025-03-01  RetailCo         51.4       5.066399
158 2025-04-01  RetailCo         51.9       5.080556
196 2025-02-01  TechCorp        200.2      14.685868
197 2025-03-01  TechCorp        197.8      14.615507
198 2025-04-01  TechCorp        202.5      14.776849
```

- **Profitability and performance are not uniformly distributed**, indicating clustering captures more than just financial health — it reveals behavioral archetypes useful for strategic segmentation.

```
Company
TechCorp     173.405000
FinServ      122.027500
HealthInc     91.166667
EnergyCo      72.344444
Name: Stock_Price, dtype: float64
```

From a decision-making standpoint, these insights can inform investment prioritization, risk assessments, and even internal auditing strategies. Moreover, the dashboard enables **dynamic filtering by company, date, and cluster**, making the analysis accessible to non-technical stakeholders.

---

4. Conclusion: Strategic Advantage Through Data Mining

This project demonstrates how **data science, when properly structured and visualized**, can bring clarity to financial decision-making. This kind of analysis can guide investment, risk management, and operational focus.