Customer Relationship Management Solution Development

The CRM solution integrates fragmented customer data, enabling stakeholders to make data-driven decisions. This document outlines the approach for developing a Customer Relationship Management (CRM) application using Palantir Foundry. The solution centralises customer interactions, stores related information, and generates actionable insights.

The project, developed as part of a job application, leverages Palantir Foundry to manage and analyse CRM data. The dataset, provided in an Excel spreadsheet with uncleaned data on companies and people, serves as the foundation. The goal is to create a CRM system that manages customer and company data, tracks interactions, leads, and opportunities, and generates insights through predictive analytics and dashboards.

1. Approach

1.1 Understanding the Problem

The CRM application is intended to:

- Centralise customer and company data.
- Track interactions, leads, and opportunities.
- Provide insights and analytics for decision-making.
- Integrate the data sources (People and Companies) to model relationships and generate meaningful metrics.
- Ensure flexibility, scalability, and the ability to handle both operational and analytical requirements.

1.2 Best Practices Adopted

- **Separation of Fields**: Atomic data storage, such as separating business type from company name to avoid redundancy.
- **Data Consistency**: Standardisation of fields like **Revenue**, **Business_Structure**, and **Contact Information** (e.g., parsing and standardising customer names and addresses).
- **Flexibility**: Enabling dynamic and efficient queries for analysis.
- **Data Integrity**: Ensuring clean and reliable data through constraints, validation rules, and relationships. Techniques like **master data management** or **partial conformity** during downstream ETL processing are used to ensure data consistency.

1.3 Tool Utilisation

Extensive use of **Palantir Foundry's** tools to develop, transform, and visualise data:

- **Code Workbook**: For data transformations, SQL queries, and Python scripting to clean and analyse `Cleaned_People` and `Cleaned_Companies` datasets.
- **Data Lineage**: To trace data transformations, ensure transparency, and troubleshoot workflows.
- **Pipeline Builder**: For designing and orchestrating scalable, reusable data pipelines.
- **Workshop**: For collaborative workflow design, model refinement, and solution iteration.
- **Notepad**: For documenting observations, compiling analysis findings, and preparing the presentation report.

2. Solution Design

2.1 Vision Document

- **Objective**: Build a scalable CRM solution that integrates datasets, extracts actionable insights, and supports predictive analysis.
- **Stakeholders**: Sales teams, data analysts, and leadership.
- **Key Capabilities**:
  - Data integration from various sources.
  - Customer segmentation and profiling.
  - Predictive analysis for sales forecasting.
  - Visualisation dashboards for actionable insights.
  - Interactive user interface for exploring data.

2.2 Data Modelling Strategy

- **Normalisation**: Clean and standardise data to ensure consistency, especially customer information (address, name parsing, etc.).
- **Bridge Tables**: Use of bridge tables for sparse attributes and multiple customer contacts (e.g., multiple addresses or customer relationships with a company).

- **Fact and Dimension Tables**: Implement fact tables for capturing transactional data (sales, interactions) and dimension tables to capture descriptive attributes (customer profiles, business structure, locations).
- **Effective Date Management**: Timespan fact tables to track effective and expiration dates for customer data, helping capture historical changes and analyse trends over time.
- **Low Cardinality Attributes**: Handling attributes with low cardinality, such as satisfaction or behaviour scores, using appropriate dimensional modelling techniques (outriggers, step dimensions).

2.3 Operational and Analytical Requirements

CRM systems must address both operational and analytic needs:

- **Operational**: Handling real-time interactions, customer inquiries, and updates to customer details.
- **Analytical**: Analysing historical data, identifying trends, and predicting future behaviours (e.g., using **predictive modelling** for sales forecasting).

2.4 Handling Dirty Data

- The uncleaned data (names, addresses) undergoes **parsing** and **standardisation** to ensure consistency and accuracy. For example, abbreviations like "Rd" are standardised to "Road," and postal codes are verified against their respective states.

3. Actionable Insights for CRM Based on Employee Data:

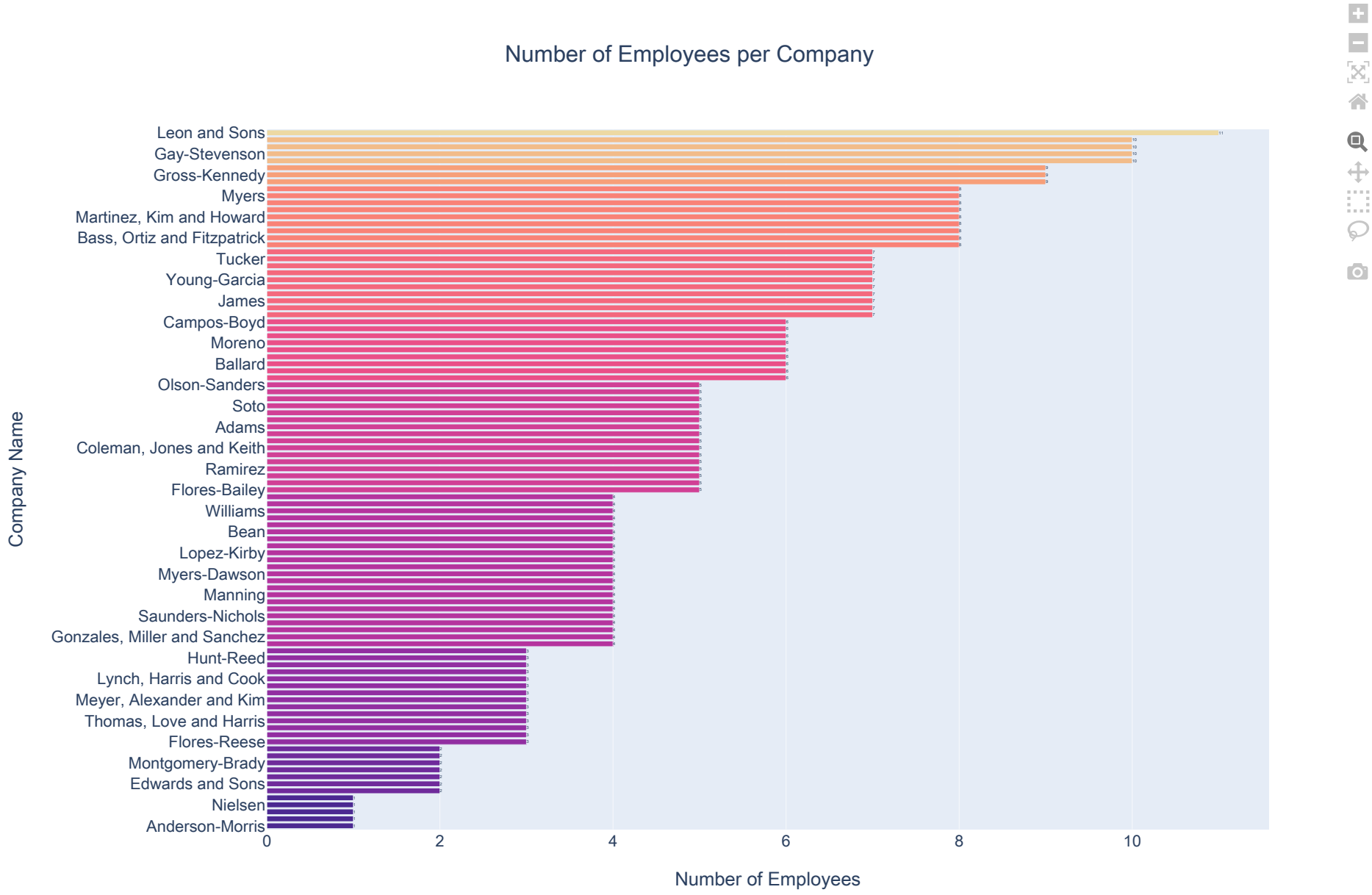
1. **Focus on Small Businesses (1–5 Employees)**
    a. The majority of companies are small. Prioritise **simple, lightweight CRMs** with automation, quick setup, and affordable pricing tiers.
2. **Support Sole Traders**
    a. Many companies have just 1 employee. Include features such as **automated follow-ups**, **reminders**, and **pre-built workflows** tailored for individual users.
3. **Collaboration for Medium-Sized Companies (6–10 Employees)**
    a. Introduce **role-based permissions**, **shared pipelines**, and team performance tracking to support growing businesses.
4. **Segment Leads by Company Size**
    a. Use employee counts to prioritise leads (e.g., larger companies = higher-value deals). Incorporate **customised outreach strategies**.

5. **Scalable Features for Growth**
    a. Provide advanced analytics, multi-contact tracking, and integrations (e.g., ERP, accounting tools) for companies with 10+ employees.
6. **Industry-Specific Templates**
    a. Offer pre-built workflows and fields for common industries (e.g., consultancy, retail). Enable customisation for niche markets.

Key CRM Features to Include:

- **Lead/Customer Management**: Track companies, roles, and interactions.
- **Pipeline Management**: Custom deal stages with prioritisation.
- **Automation Tools**: Follow-up reminders, email campaigns.
- **Reporting**: Segmentation by company size and deal tracking.
- **Scalability**: Affordable for small teams, expandable for larger ones.
- **Integration**: Connect with email, marketing, and financial tools.

This ensures the CRM is well-suited to micro and mid-sized businesses, fostering engagement and growth.

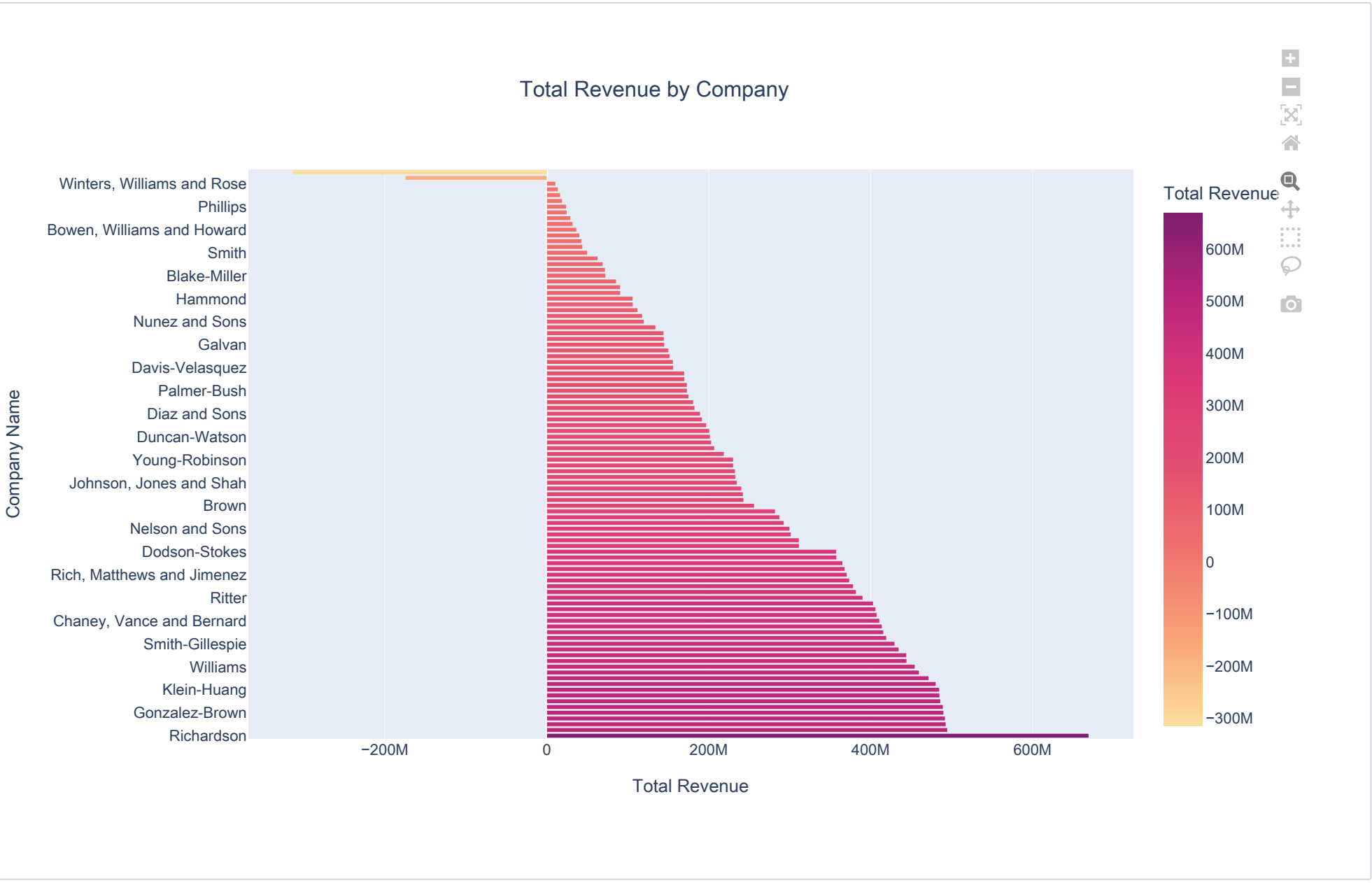

Number of Employees per Company
Company Name
Leon and Sons
Gay-Stevenson
Gross-Kennedy
Myers
Martinez, Kim and Howard
Bass, Ortiz and Fitzpatrick
Tucker
Young-Garcia
James
Campos-Boyd
Moreno
Ballard
Olson-Sanders
Soto
Adams
Coleman, Jones and Keith
Ramirez
Flores-Bailey
Williams
Bean
Lopez-Kirby
Myers-Dawson
Manning
Saunders-Nichols
Gonzales, Miller and Sanchez
Hunt-Reed
Lynch, Harris and Cook
Meyer, Alexander and Kim
Thomas, Love and Harris
Flores-Reese
Montgomery-Brady
Edwards and Sons
Nielsen
Anderson-Morris
0
2
4
6
8
10
Number of Employees

4. Insight: Revenue vs. Company Size

1. **High Revenue Companies:**
    a. Companies like *Richardson*, *Collins, Coleman and Ortega*, and *Wiley Group* generate revenues exceeding $490M. They likely have a large number of employees, as high revenues often correlate with larger workforces. These companies are good candidates for CRM solutions focused on **streamlining employee-customer interactions**.
2. **Mid-Tier Revenue Companies:**
    a. Companies such as *Johnson*, *Smith-Gillespie*, and *Maddox* generate $300M–$450M. These may have medium-sized workforces. CRM features prioritising **lead generation, customer segmentation, and opportunity tracking** would provide significant value here.
3. **Low Revenue or Struggling Companies:**
    a. Firms like *Moore, Johnson and Sanchez* and *Fisher-Stevenson* are showing negative revenue. These companies may benefit from **cost-effective CRM options** that enhance productivity and customer retention without requiring extensive resources.
4. **Small Revenue Companies:**
    a. Firms such as *Phillips*, *Clark*, and *Obrien-Parker* generate less than $25M. Likely small businesses, they may require lightweight CRM systems focusing on **contact management** and **pipeline visualisation**.

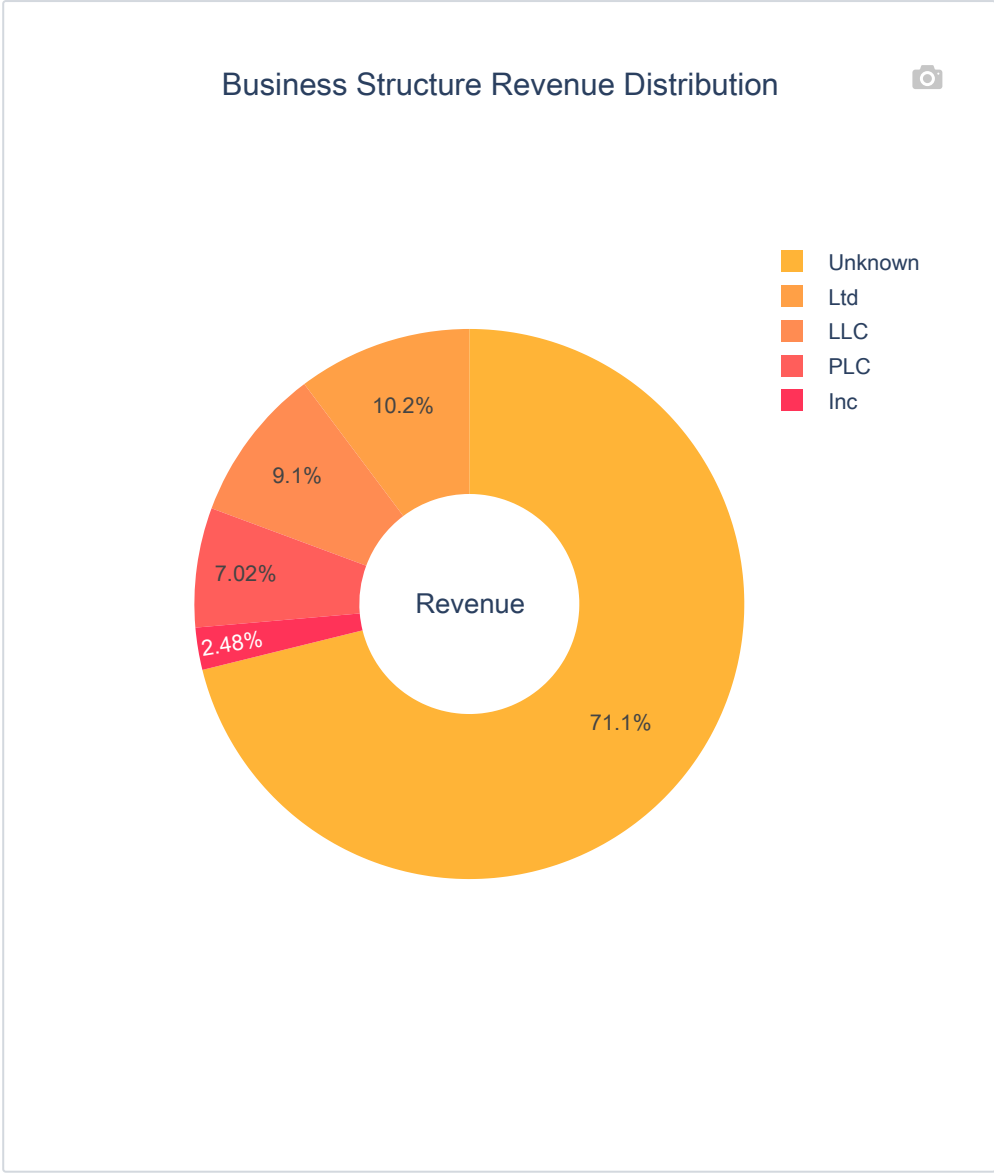
Strategic Recommendation:

Prioritise targeting **mid-to-high revenue companies**, as they are likely to have the workforce size and financial capability to adopt more sophisticated CRM solutions. Customise CRM offerings based on their revenue tier and size to maximise adoption and value. For struggling companies, introduce free or low-cost versions to build trust and foster future partnerships.



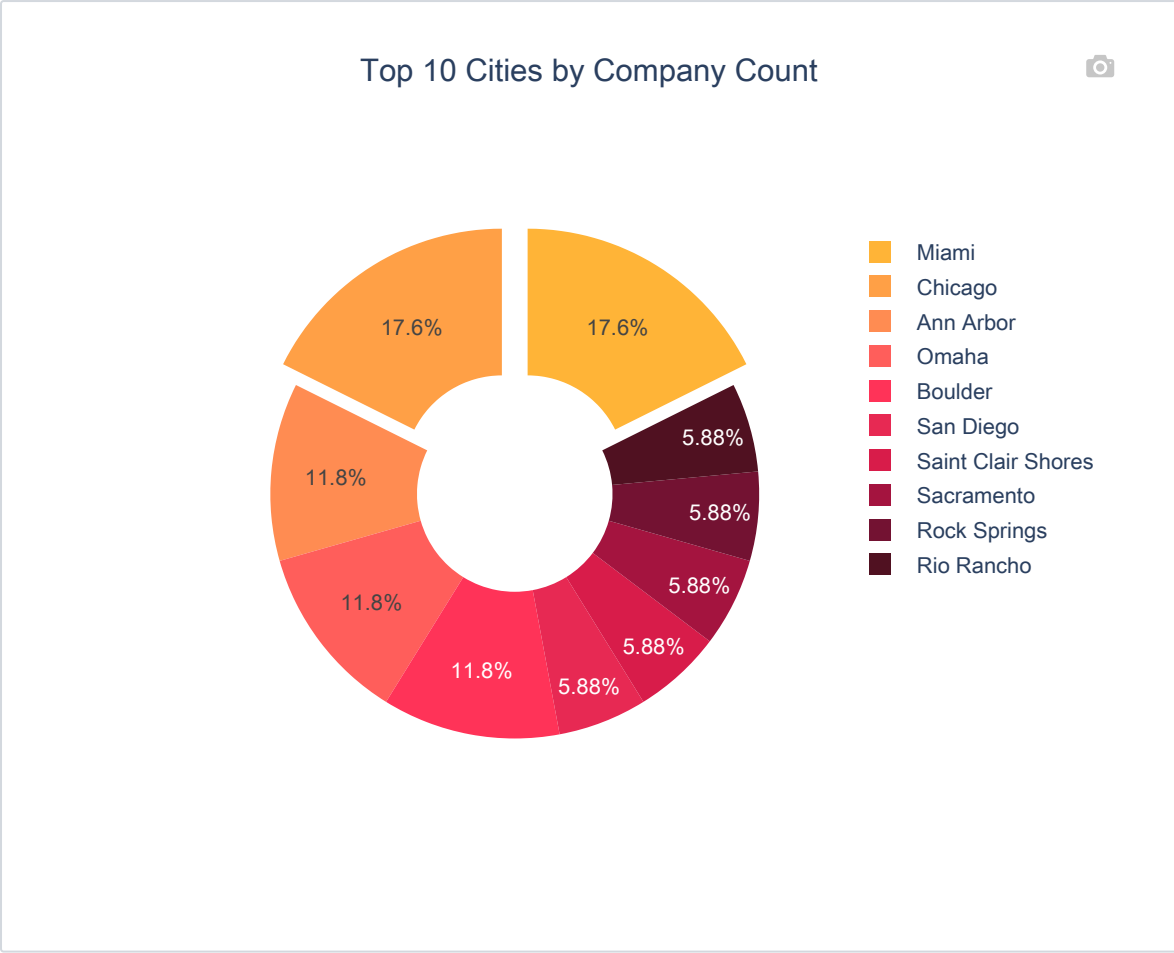
5. Actionable Insights for Business Structure Revenue Distribution

1. **Focus on "Unknown" Companies**
   The majority of revenue ($16.86 billion, 71.15%) comes from companies with an "Unknown" business structure. This indicates incomplete data, which could hinder targeting efforts. Prioritise identifying and classifying these companies to enable more tailored engagement strategies.
2. **Target High-Revenue Segments**
   a. **Ltd** companies contribute the second-largest revenue ($2.43 billion, 10.25%). These companies may offer substantial growth opportunities for sales teams.
   b. Similarly, **LLC** companies generate $2.16 billion (9.10%) in revenue and should be a key focus.
3. **Lower Priority Segments**
   a. **Inc** companies contribute only $588.44 million (2.48%). Consider reallocating resources to higher-performing segments unless there is strategic interest in this business structure.
4. **Customised CRM Capabilities**
   a. Include segmentation tools by **Business Structure** and **Revenue Bands** to help sales teams prioritise high-revenue clients.
   b. Implement data enrichment workflows to address missing information for "Unknown" companies.
5. **Revenue-Driven Outreach**
   Focus outreach efforts on high-revenue companies, especially those categorised as **Ltd** and **LLC**, while improving data quality for "Unknown" companies to unlock untapped revenue potential.



6. Actionable Insights for Top 10 Companies by City:

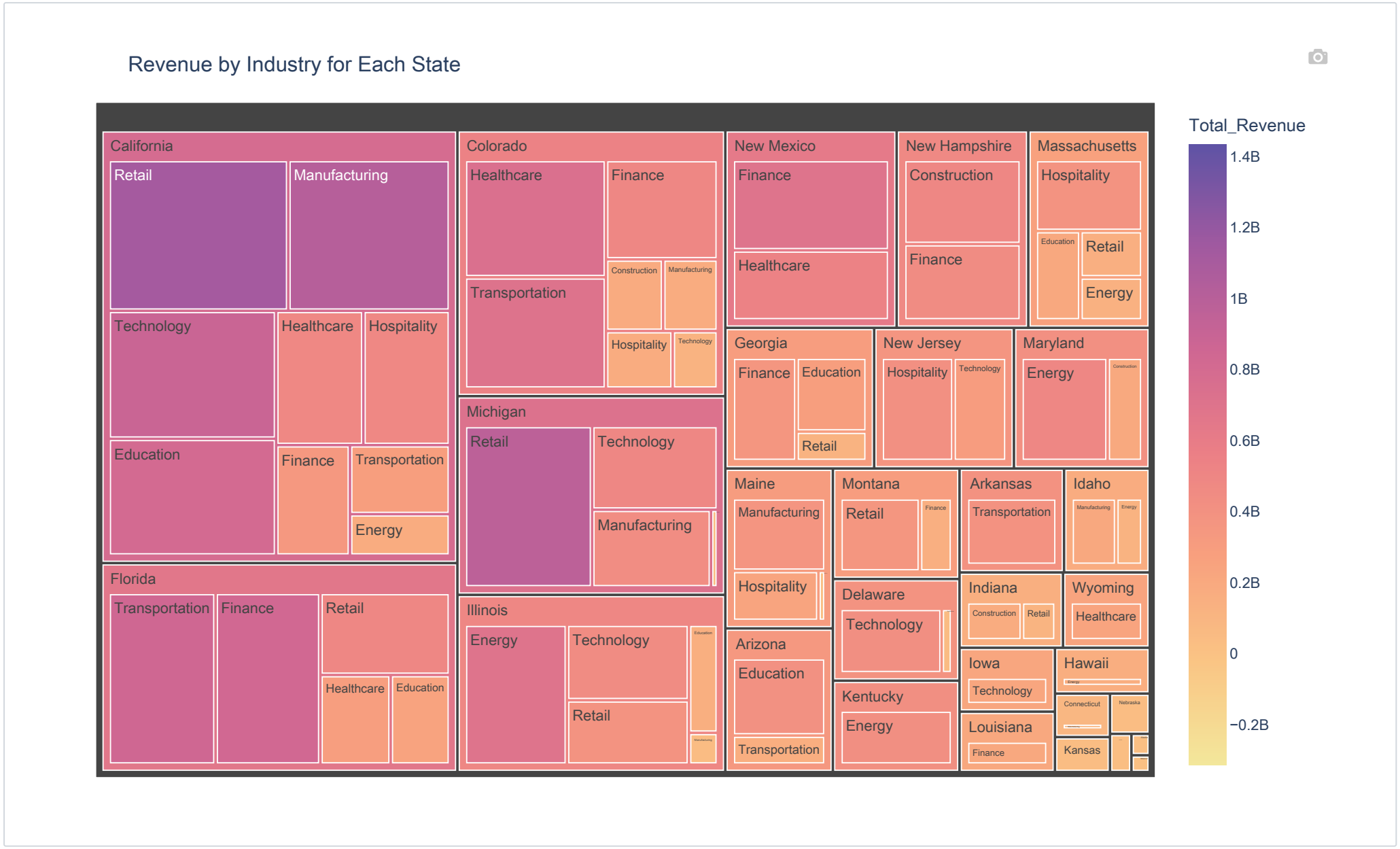
Based on the number of employees per company across various cities, here are key insights:



7. Actionable Insights for Revenue Industries by State

- **Miami, Chicago**: Both cities account for 17.65% of the companies in the dataset, with each having three companies. These cities show a higher concentration of businesses, indicating a potentially larger market for customer acquisition and sales.
- **Ann Arbor, Omaha, Boulder**: These cities have two companies each, making up 11.76% of the total. While still significant, they represent a smaller proportion compared to Miami and Chicago.
- **San Diego, Saint Clair Shores, Sacramento, Rock Springs, Rio Rancho**: Each of these cities has one company, making up 5.88% of the total. These cities represent niche markets with fewer businesses, which could mean less competition but potentially lower revenue opportunities.

This distribution suggests a stronger sales opportunity in **Miami** and **Chicago** due to the higher number of companies, with **Ann Arbor**, **Omaha**, and **Boulder** offering moderate opportunities. The smaller cities could be targeted for specialised services or products, given their limited competition.



Based on the data provided, here are some actionable insights:

1. **Top Performing Industries by Revenue:**

a. **Retail:** California leads with $1.1 billion in revenue, followed by Michigan at $962 million. Retail is a strong industry across various states, including Florida and Montana.

b. **Healthcare:** California, Colorado, and New Mexico are leading with significant revenue. California alone contributes over $480 million.

c. **Energy:** Kentucky, Maryland, and Illinois lead with over $400 million, indicating energy is a major sector, particularly in these states.

2. **States with Significant Revenue Contributions:**

   a. **California:** Dominates with high revenue across multiple industries, including Retail ($1.1 billion), Manufacturing ($1 billion), Technology ($879 million), and Education ($803 million).

   b. **Michigan and Florida** are also top contributors, especially in Retail, Finance, and Healthcare.

3. **States with Negative or Low Revenue:**

   a. **Minnesota:** Both Education and Hospitality have significant negative revenue, indicating potential financial losses.

   b. **Kansas and Missouri:** These states show relatively low revenue across multiple sectors like Hospitality, Transportation, and Construction.

4. **Industries to Focus On:**

   a. **Retail and Healthcare:** These industries have consistently high revenues, particularly in states like California, Michigan, and Colorado.

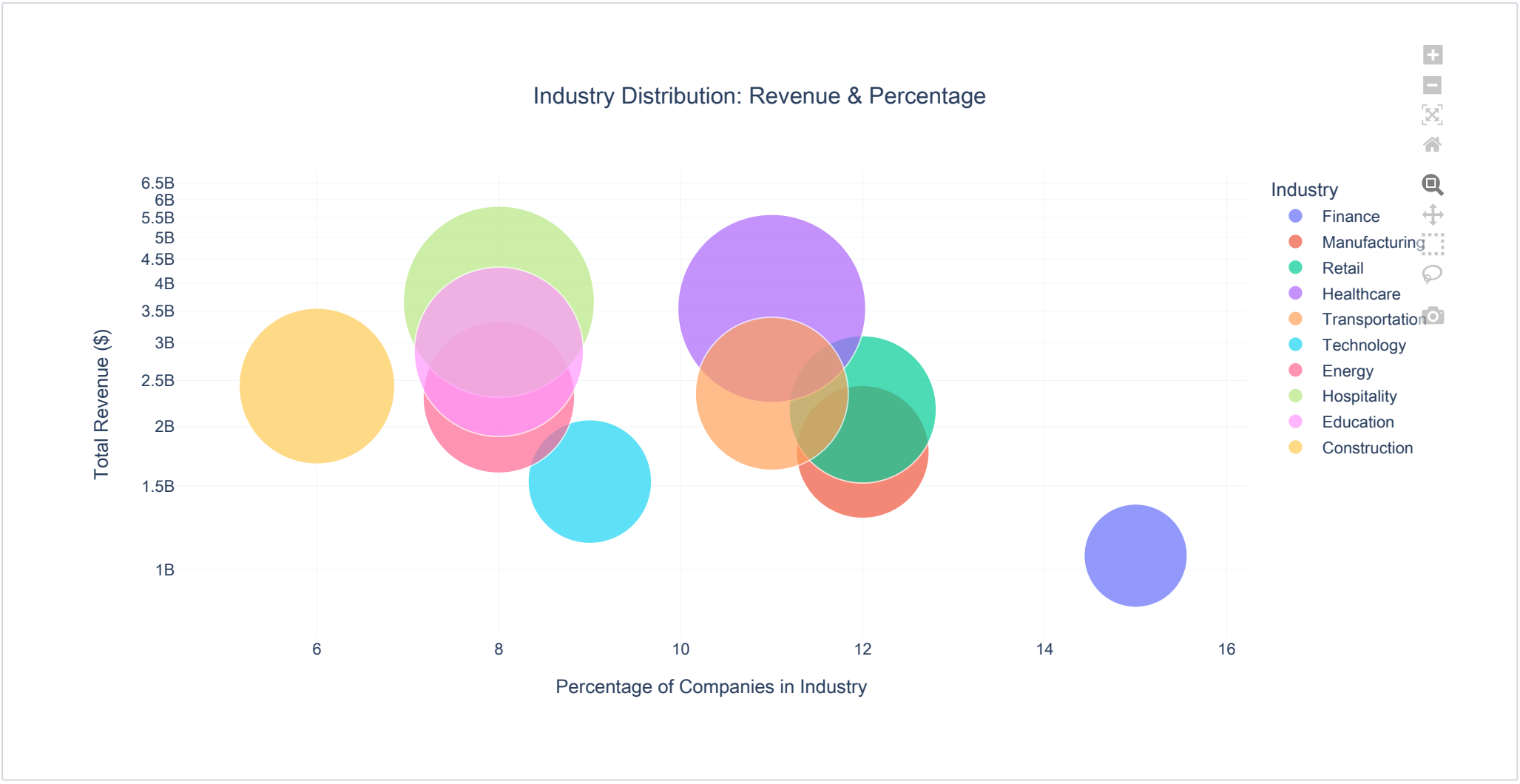
   b. **Construction:** States like New Hampshire and Indiana show significant revenues, presenting potential growth areas.

5. **Opportunity for Expansion:**

   a. **Energy and Technology:** States like Illinois, Maryland, and Delaware have strong performance in these sectors, suggesting potential for business opportunities, particularly in sustainable energy and tech innovations.

In summary, focusing on retail, healthcare, and energy industries, particularly in California and Michigan, presents the most significant opportunities for growth. Additionally, addressing negative revenue trends in Minnesota and Kansas could mitigate risks.

8. Actionable Insights for Industry Distribution: Revenue & Percentage



Based on the data, the **Finance** industry contributes the largest share of revenue at $1.07 billion, followed by **Healthcare** and **Hospitality** with $3.54 billion and $3.66 billion, respectively. The **Retail** and **Manufacturing** industries each generate around $2.17 billion and $1.77 billion, respectively.

When considering the number of employees per company, larger revenue-generating sectors like **Healthcare**, **Retail**, and **Energy** likely have a significant workforce, which could indicate a higher need for CRM capabilities to manage customer interactions efficiently, streamline sales, and optimise customer relationships.

Key Insight: Companies in high-revenue industries, especially **Healthcare** and **Retail**, would benefit from advanced CRM features like lead management, opportunity tracking, and analytics to support their large customer bases and complex sales processes.

9. Storytelling with Data

9.1 Narrative Style

- **Setup**: The datasets revealed a wide range of trends in customer information, revenue generation, and business structure, but were underutilised without proper integration and analysis.
- **Conflict**: The data was fragmented, lacked predictive capabilities, and wasn't easily accessible for real-time decision-making.
- **Resolution**: The CRM solution integrates these datasets into a single, unified platform, providing predictive insights, actionable dashboards, and tools for informed decision-making.