Tableau Data Linking Case Study

### Overview

This case study demonstrates how to load, cleanse, and link datasets from various sources in Tableau to enable multi-dimensional analysis. It includes working with a large ZIP-code business dataset, a population estimate dataset, and a state abbreviation mapper to create meaningful tables and visualizations.

### Data Files Used

1. **zbp21detail.txt** (2.9M rows): ZIP Code Industry Detail File (2021 business data)
2. **population.csv**: Cleaned population estimates by state (2020–2022)
3. **abbreviations.xlsx**: Maps full state names to abbreviations

### Key Learning Objectives

* Set relationships across files from different formats
* Adjust cardinality and ordinality in Tableau
* Clean and transform fields (e.g., remove characters, rename columns)
* Create calculated fields
* Validate joins using tables
* Generate scatterplot visualizations to explore insights

### Tables Created

* **Table 1:** State Abbreviation vs. Total Establishments (from zbp21detail.txt)

A screenshot of a computer

AI-generated content may be incorrect.

* **Table 2:** Full State Name vs. Establishments (joins through abbreviation mapping)

A screenshot of a computer

AI-generated content may be incorrect.

* **Table 3:** Calculated businesses per 100K population ("Est / Y2022 \* 100000")

A screenshot of a computer

AI-generated content may be incorrect.

### Visualizations

* **Scatterplot 1: Orgs by Population**
  + Shows correlation between state population and business count
  + X-axis: Population (Y2022), Y-axis: Establishments

A screenshot of a computer screen

AI-generated content may be incorrect.

* **Scatterplot 2: Tiny Orgs vs. Population**
  + Uses n<5 as proxy for entrepreneurship
  + X-axis: Population, Y-axis: Establishments with <5 employees

A screenshot of a computer

AI-generated content may be incorrect.

### Insights Uncovered

* States with larger populations generally have more business establishments, confirming expected trends.
* Alaska, despite its small population, has a disproportionately high number of businesses per capita, suggesting either more distributed small businesses or reporting nuances.
* Some small-population states show relatively high levels of tiny organizations (n<5), indicating potential entrepreneurial activity hotspots.
* California, while having the highest raw number of businesses, does not dominate in terms of per capita or micro-business.

### Conclusion

By joining and cleansing data from different formats, this project validates data integrity and uncovers insights about business distribution and entrepreneurship across U.S. states.