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

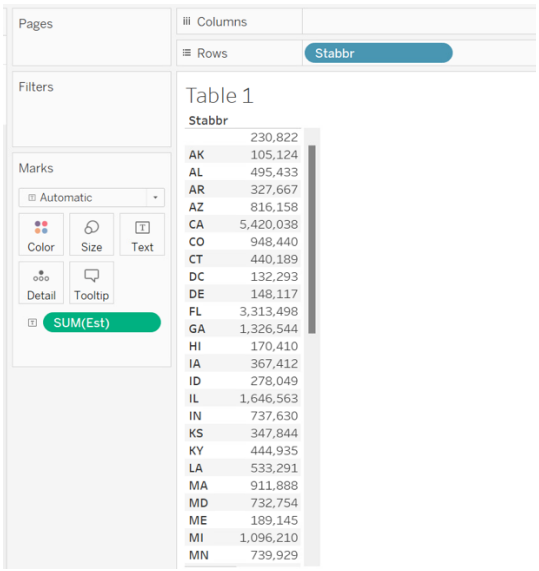
1. **zbp21detail.txt** (2.9M rows): ZIP Code Industry Detail File (2021 business data)
2. **population.csv**: Cleaned population estimates by state (2020–2022)
3. **abbreviations.xlsx**: Maps full state names to abbreviations

Key Learning Objectives

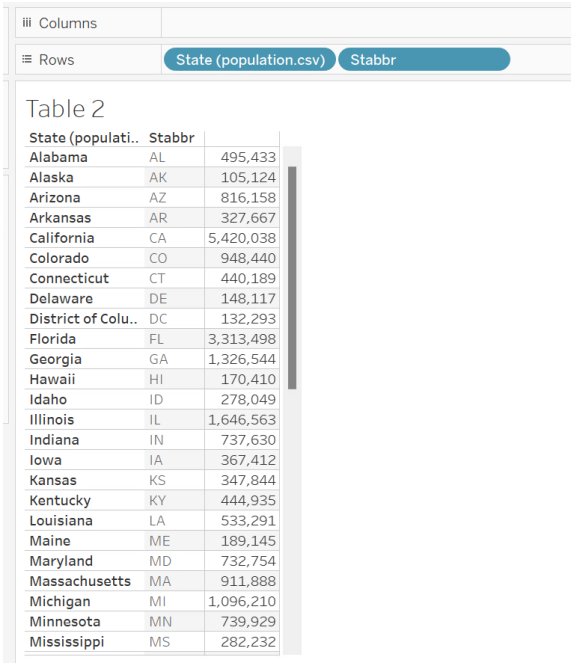
- Set relationships across files from different formats
- Adjust cardinality and ordinality in Tableau
- Clean and transform fields (e.g., remove characters, rename columns)
- Create calculated fields
- Validate joins using tables
- Generate scatterplot visualizations to explore insights

Tables Created

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

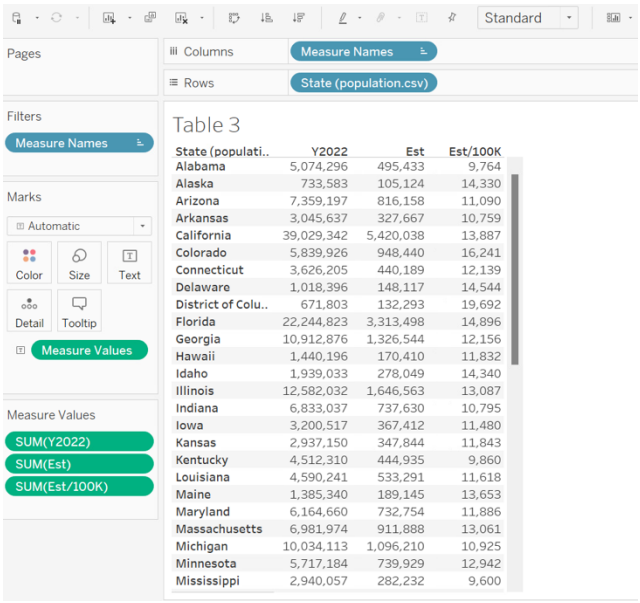


- **Table 2:** Full State Name vs. Establishments (joins through abbreviation mapping)



State (populati..	Stabbr	
Alabama	AL	495,433
Alaska	AK	105,124
Arizona	AZ	816,158
Arkansas	AR	327,667
California	CA	5,420,038
Colorado	CO	948,440
Connecticut	CT	440,189
Delaware	DE	148,117
District of Colu..	DC	132,293
Florida	FL	3,313,498
Georgia	GA	1,326,544
Hawaii	HI	170,410
Idaho	ID	278,049
Illinois	IL	1,646,563
Indiana	IN	737,630
Iowa	IA	367,412
Kansas	KS	347,844
Kentucky	KY	444,935
Louisiana	LA	533,291
Maine	ME	189,145
Maryland	MD	732,754
Massachusetts	MA	911,888
Michigan	MI	1,096,210
Minnesota	MN	739,929
Mississippi	MS	282,232

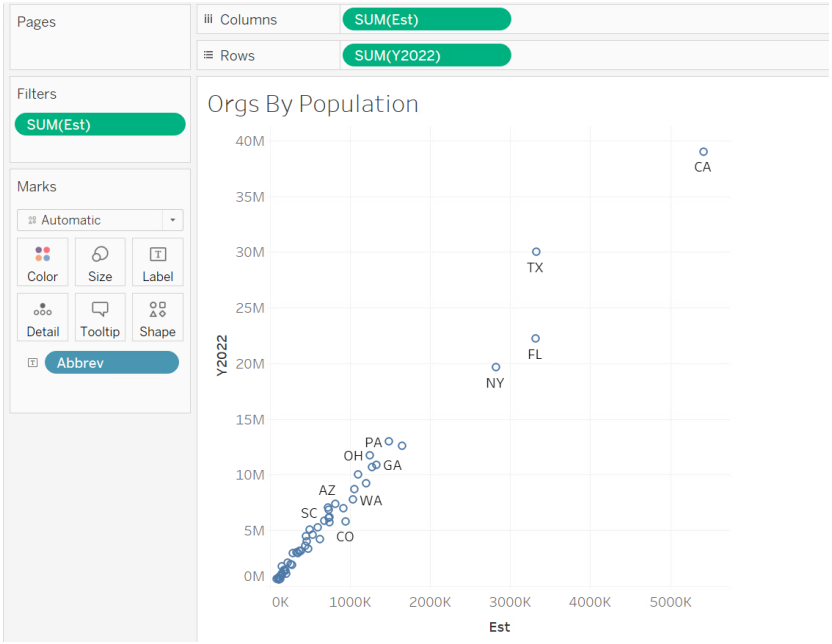
- **Table 3:** Calculated businesses per 100K population ("Est / Y2022 * 100000")



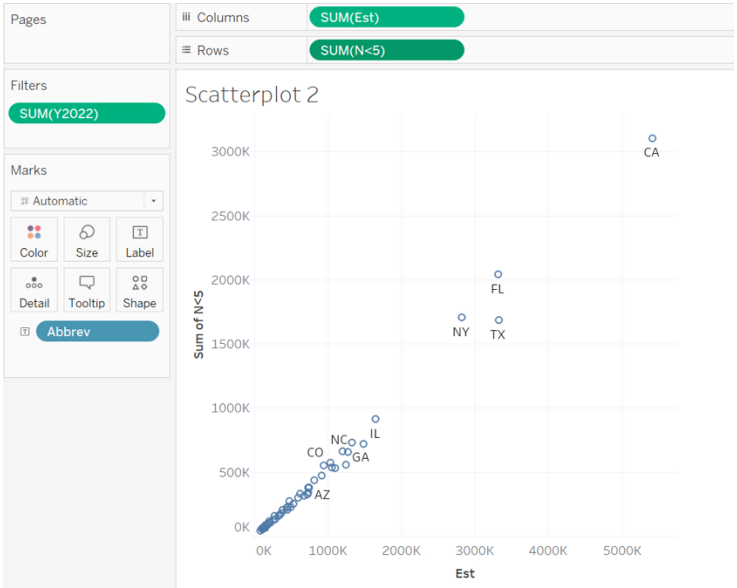
State (populati..	Y2022	Est	Est/100K
Alabama	5,074,296	495,433	9,764
Alaska	733,583	105,124	14,330
Arizona	7,359,197	816,158	11,090
Arkansas	3,045,637	327,667	10,759
California	39,029,342	5,420,038	13,887
Colorado	5,839,926	948,440	16,241
Connecticut	3,626,205	440,189	12,139
Delaware	1,018,396	148,117	14,544
District of Colu..	671,803	132,293	19,692
Florida	22,244,823	3,313,498	14,896
Georgia	10,912,876	1,326,544	12,156
Hawaii	1,440,196	170,410	11,832
Idaho	1,939,033	278,049	14,340
Illinois	12,582,032	1,646,563	13,087
Indiana	6,833,037	737,630	10,795
Iowa	3,200,517	367,412	11,480
Kansas	2,937,150	347,844	11,843
Kentucky	4,512,310	444,935	9,860
Louisiana	4,590,241	533,291	11,618
Maine	1,385,340	189,145	13,653
Maryland	6,164,660	732,754	11,886
Massachusetts	6,981,974	911,888	13,061
Michigan	10,034,113	1,096,210	10,925
Minnesota	5,717,184	739,929	12,942
Mississippi	2,940,057	282,232	9,600

Visualizations

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



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



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.