ROBERT H. SMITH
SCHOOL OF BUSINESS

Amtrak Consulting Project
TerpTrak Solutions

Palak Kishore, Ashna Sandoze, Manisha Ahuja, Maitreyee Tiwari
Professor Lee
BUDT703_0506
07 December 2024

Data Processing:

We started with a summarized dataset provided by our professor in an Excel file. The initial step was to review the data structure and determine how to rearrange it to fit our desired relational schema. This process involved reformatting, reorganizing, and cleaning the data to ensure it was consistent and compatible with the schema.

After preparing the data, we used SQL Server's "Import Flat File" task to load the Excel file into our database. This step made the data accessible within the SQL Server environment, enabling us to run queries, perform analyses, and integrate it with other datasets. By following these steps, we ensured the data was properly structured and ready for use in our project.

Analysis of Ridership Growth and Budget/Procurement Allocations:

Question #1:
Which stations have the lowest ridership growth, and how does this compare to the stations total procurement and budget allocations?

Our Process
Using SQL, we created a table that analyzed the ridership growth, procurement amounts, and budget allocations for the stations. This query was overall able to calculate growth metrics and give us a clearer idea of which stations had a larger growth in ridership. Shown below are our query results on the left is the stations with the lowest ridership growth and on the right is the stations with the highest ridership growth:

	State Name	Station Name	Ridership Growth	Total Procurement	Total Budget
1	South Carolina	Camden	1262	933858	3496000.00
2	Arizona	Winslow	1536	150766	7804000.00
3	Colorado	La Junta	1877	1920809	3479000.00
4	Georgia	Gainesville	2321	57046457	1941000.00
5	Nevada	Elko	2365	397634	460000.00
6	Louisiana	Slidell	2402	120349	1675000.00
7	New York	Amsterdam	3092	253988	274000.00
8	South Carolina	Greenville	4296	189848	4103000.00
9	Pennsylvania	Erie	6235	85863993	300000.00
10	North Dakota	Fargo	7035	67576795	2882000.00

	State Name	Station Name	Ridership Growth	Total Procurement	Total Budget
1	Illinois	Chicago-Union Station	1385923	317412916	9256000.00
2	California	Los Angeles-Union Station	533826	78792575	237000.00
3	Rhode Island	Providence	405704	1008871	302000.00
4	Delaware	Wilmington	397454	58247777	758000.00
5	California	Emeryville	307431	157643	528000.00
6	Pennsylvania	Lancaster	200967	2233340	606000.00
7	California	Oakland-Jack London S...	101474	985132	824000.00
8	California	Richmond	94123	2886026	3000000.00
9	Missouri	Kansas City-Union Station	61583	15797971	8374000.00
10	New York	Syracuse-Regional Tran...	60013	1518208	0.00

We then were able to visualize our findings in Tableau as shown below:

The image on the left shows the stations that have the lowest ridership growth and how it compares to the stations procurement and budget allocations. The image on the right shows the stations that have the highest ridership growth and how it compares to the stations procurement and budget allocations.

Our Findings

In our findings, we found a weak correlation between higher budgets and higher ridership growth. For example, stations such as Chicago Union and Kansas Union Station demonstrate a positive correlation between high budgets and an increase in ridership growth. Although there was a correlation, we found that this was not enough evidence to prove that the higher budget would increase ridership growth. For example, stations such as Camden and Winslow there is low growth despite there being a higher budget. We also noticed that Gainesville, Fargo, and Erie stations show high procurement amounts but low ridership growth, this could be as a result of inefficiencies or potential mismanagement of funds.

Our Recommendations:

Given our analysis, we would recommend the following:

1. Develop a targeted plan – this would include during an analysis on where the specific issues are and improve the specific issues to ensure that budget is being utilized efficiently.
2. Align budget distribution with stations demonstrating potential for ridership growth while addressing inefficiencies at high-budget, low-growth stations.
3. Establish robust monitoring systems to track budget utilization and performance metrics, ensuring funds are effectively contributing to ridership growth.

Analysis of Ridership vs. Rewards Participation by State:

Question #2:

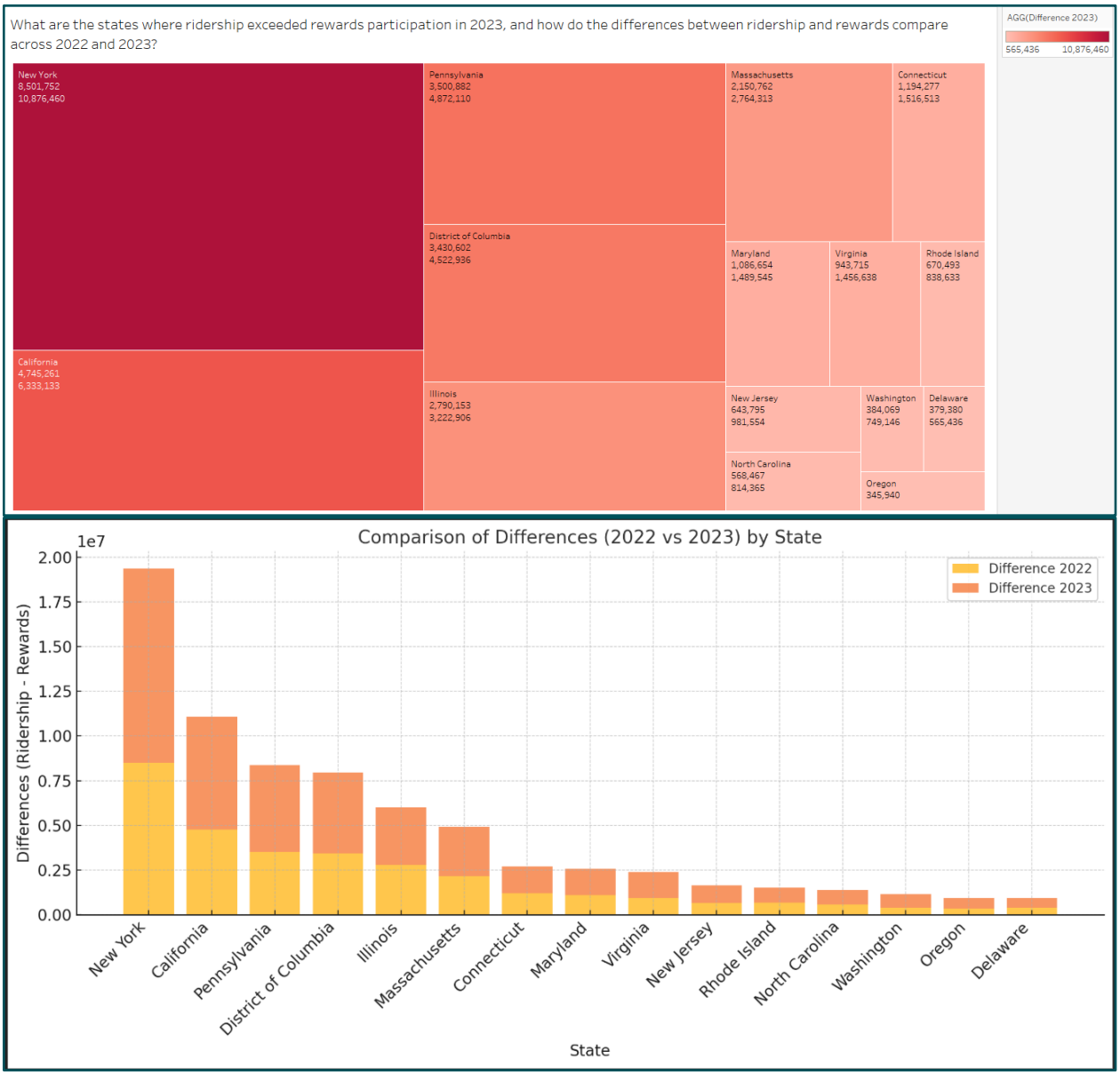
What are the states where ridership exceeded rewards participation in 2023, and how do the differences between ridership and rewards compare across 2022 and 2023?

Our Process:

Using SQL, we analyzed the relationship between reward participation and ridership across all states, specifically for 2022 and 2023. We created a table that highlighted the states where ridership was greater than rewards participation in 2023 as shown below:

	State Name	Ridership 2022	Rewards 2022	Difference 2022	Ridership 2023	Rewards 2023	Difference 2023
1	New York	9888379	1386627	8501752	12454331	1577871	10876460
2	California	6420985	1675724	4745261	8207740	1874607	6333133
3	Pennsylvania	4182216	681334	3500882	5647603	775493	4872110
4	District of Columbia	3631677	201075	3430602	4751405	228469	4522936
5	Illinois	3389218	599065	2790153	3898203	675297	3222906
6	Massachusetts	2679581	528819	2150762	3375479	611166	2764313
7	Connecticut	1448437	254160	1194277	1808203	291690	1516513
8	Maryland	1575986	489332	1086654	2053928	564383	1489545
9	Virginia	1578326	634611	943715	2194700	738062	1456638
10	New Jersey	1173868	530073	643795	1576582	595028	981554
11	Rhode Island	756776	86283	670493	939329	100696	838633
12	North Carolina	863006	294539	568467	1156546	342181	814365
13	Washington	711588	327519	384069	1120025	370879	749146
14	Oregon	546938	200998	345940	814805	228222	586583
15	Delaware	456001	76621	379380	654717	89281	565436

After gathering our data we then chose to visualize the trends in the data using Tableau, to get a better understanding of how much of a gap there is between ridership and rewards engagement. Below are our graphs from Tableau.



Our Findings

In our findings, we saw that there were larger gaps between ridership and rewards participation in states that have a higher population. For example New York and California have a large gap, this may be as a result of a large commuter population that may rely on transit systems but are not a part the reward program. We also noticed that as time went, the gap between the ridership and rewards grew as the rewards program did not increase at the same pace as ridership. Overall, we saw a huge difference between the amount of riders and rewards and saw there was room for improvement.

Our Recommendations:

Given our analysis, we would recommend the following:

1. Investing in market research and develop marketing strategies to gain awareness for the rewards system. This should include doing an analysis on current marketing efforts to pinpoint the areas that are not reaching the intended audience and then finding ways to improve the gap between ridership and the rewards participation .

2. Add incentives to joining the rewards program for frequent users. For example for every trip you purchase you can earn miles, the miles could then be redeemed to purchase future travel.
3. Physical signage at stations that Amtrak users can see as they are travelling and gain awareness of the rewards program.

Analysis of Routes with Lowest On-Time Performance (OTP)

Question #3

Which are the bottom fifteen routes with the lowest on-time performance (OTP), and what factors contribute to their low performance?

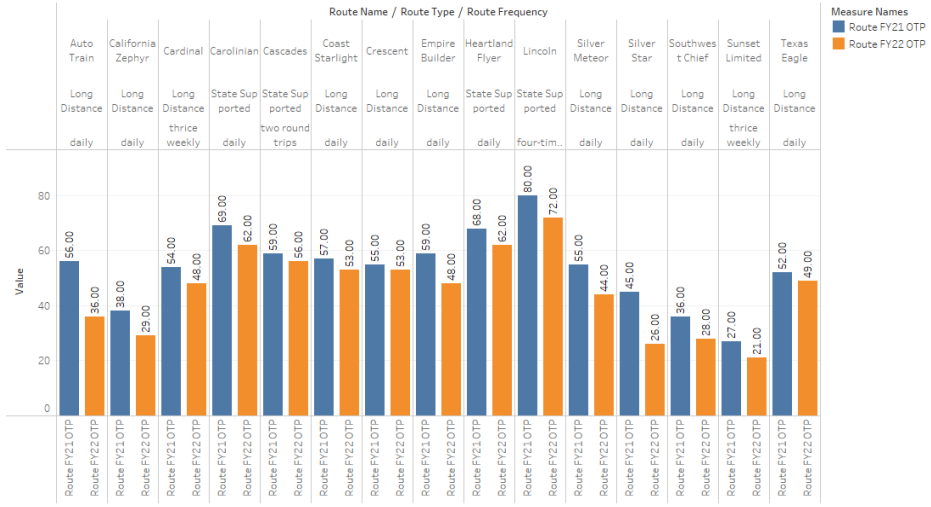
Our Process

Using SQL, we gathered OTP data for FY21-23 for all the routes to be able to track the performance over the three years. We then added other variables to reveal any other information that may be insightful to our analysis, such as route type and frequency. We were ultimately able to gather information on whether the OTP improved, declined, or stayed consistent over time. Our query results are shown below:

	Route Name	Route Type	Route Frequency	Route FY21 OTP	Route Trend (FY21-FY22)	Route FY22 OTP	Route Trend (FY22-FY23)	Route FY23 OTP	Route Trend (FY21-FY23)
1	California Zephyr	Long Distance	daily	38.00%	Decrease	29.00%	Increase	33.00%	Decrease
2	Southwest Chief	Long Distance	daily	36.00%	Decrease	28.00%	Increase	34.00%	Decrease
3	Sunset Limited	Long Distance	thrice weekly	27.00%	Decrease	21.00%	Increase	44.00%	Increase
4	Silver Star	Long Distance	daily	45.00%	Decrease	26.00%	Increase	49.00%	Increase
5	Empire Builder	Long Distance	daily	59.00%	Decrease	48.00%	Increase	51.00%	Decrease
6	Silver Meteor	Long Distance	daily	55.00%	Decrease	44.00%	Increase	53.00%	Decrease
7	Crescent	Long Distance	daily	55.00%	Decrease	53.00%	Increase	57.00%	Increase
8	Coast Starlight	Long Distance	daily	57.00%	Decrease	53.00%	Increase	58.00%	Increase
9	Cardinal	Long Distance	thrice weekly	54.00%	Decrease	48.00%	Increase	58.00%	Increase
10	Texas Eagle	Long Distance	daily	52.00%	Decrease	49.00%	Increase	60.00%	Increase
11	Carolinian	State Supported	daily	69.00%	Decrease	62.00%	Decrease	61.00%	Decrease
12	Lincoln	State Supported	four-times-daily	80.00%	Decrease	72.00%	Decrease	63.00%	Decrease
13	Heartland Flyer	State Supported	daily	68.00%	Decrease	62.00%	Increase	64.00%	Decrease
14	Cascades	State Supported	two roundtrips	59.00%	Decrease	56.00%	Increase	64.00%	Increase
15	Auto Train	Long Distance	daily	56.00%	Decrease	36.00%	Increase	66.00%	Increase

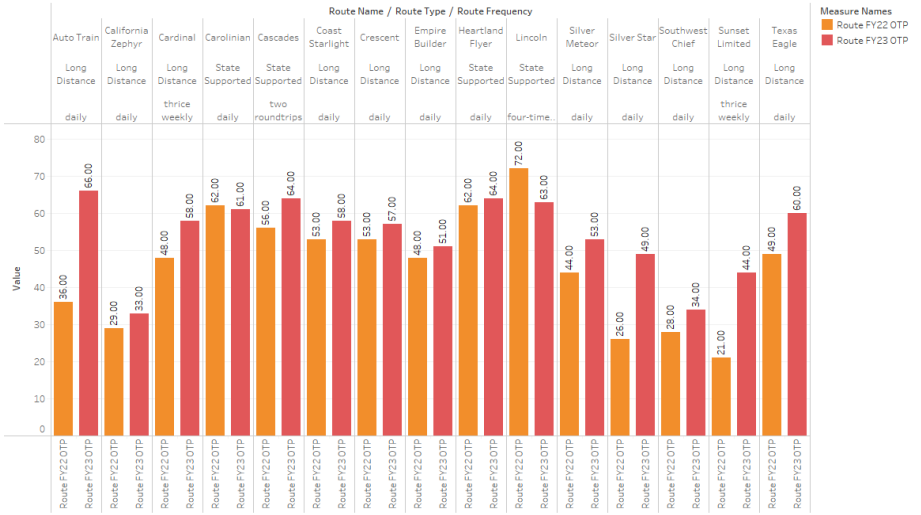
Below are our graphs from Tableau.

Which are the bottom fifteen routes with the lowest on-time performance (OTP), and what factors contribute to their low performance?

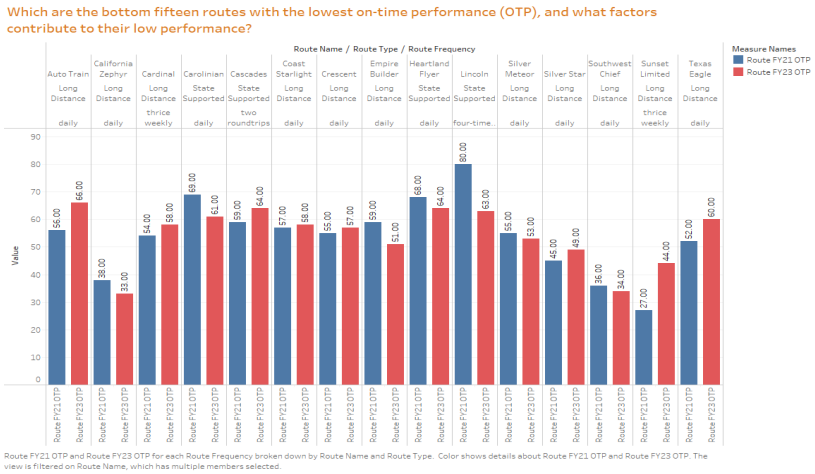


Route FY21 OTP and Route FY22 OTP for each Route Frequency broken down by Route Name and Route Type. Color shows details about Route FY21 OTP and Route FY22 OTP. The view is filtered on Route Name, which has multiple members selected.

Which are the bottom fifteen routes with the lowest on-time performance (OTP), and what factors contribute to their low performance?



Route FY22 OTP and Route FY23 OTP for each Route Frequency broken down by Route Name and Route Type. Color shows details about Route FY22 OTP and Route FY23 OTP. The view is filtered on Route Name, which has multiple members selected.



Our Findings

In our findings, we saw that long distance routes have a lower OTP, this could be a result of greater exposure in delays. For example, given that there are many stops, if the first stop is delayed, the subsequent stops will also be delayed. Since there are many stops in this route it will be on time less than a route that may have a few stops. We also saw that several routes had a decrease in OTP from the years of FY21 and FY23. Lastly we saw that routes with daily services, are being improved slowly, this could be a result of daily commuters wanting better service as they may rely on the trains to go to work

Our Recommendations:

Given our analysis, we would recommend the following:

1. Complete an analysis on what specific factors cause the delays – figuring out what the delays are caused by, whether they are environmental or caused by people would be beneficial in
2. Take a look at preventative measures – see what can be done to prevent delays and improve customer satisfaction.
3. Optimize scheduling times, ensure that there is enough buffer time to offset potential delays in routes.

Analysis of Stations with Diminishing Ridership

Question #4

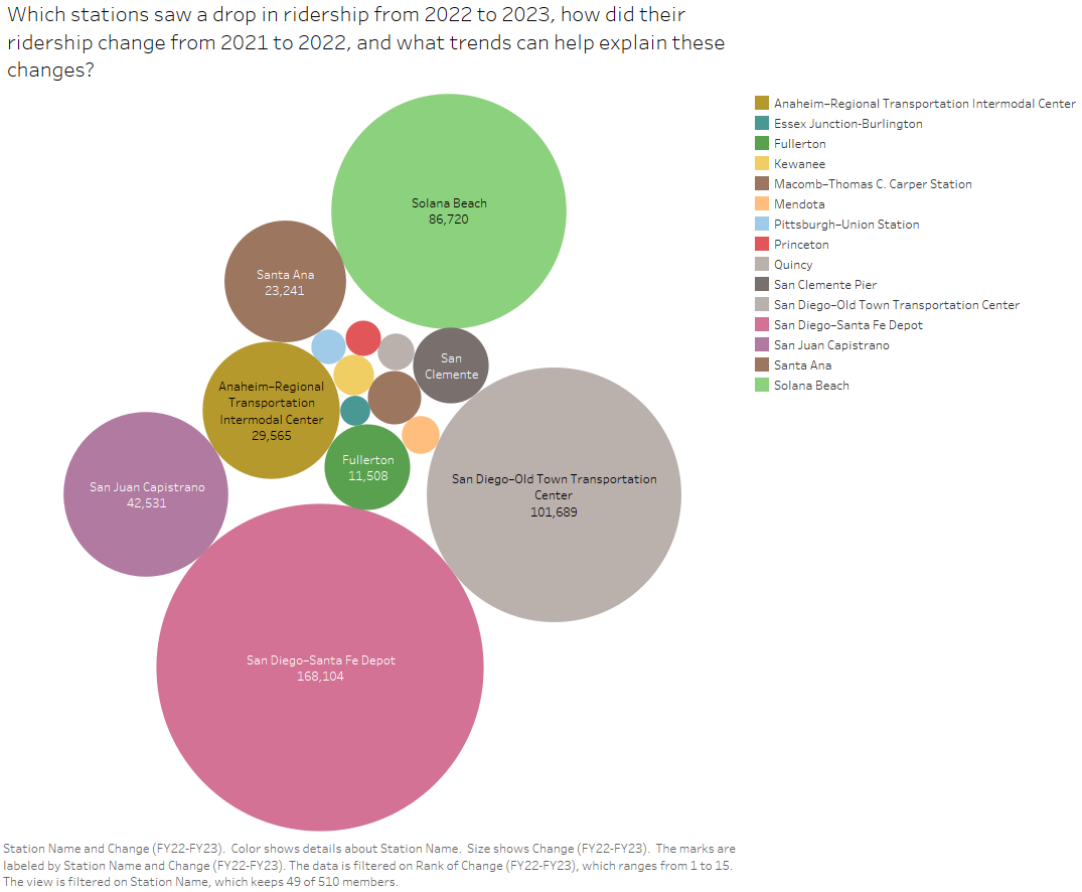
Which stations saw a drop in ridership from 2022 to 2023, how did their ridership change from 2021 to 2022, and what trends can help explain these changes?

Our Process

Using SQL we were able to gather the ridership data from FY21-FY23 for all the stations to gain insight on performance over the years. We then calculated all the ridership changes over the span on one year to uncover which years had a decrease in ridership. The following was a result of this query:

	Station Name	FY21 Ridership	Ridership Change (FY21-FY22)	Ridership Trend (FY21-FY22)	FY22 Ridership	Ridership Change (FY22-FY23)	Ridership Trend (FY22-FY23)	FY23 Ridership
1	San Diego–Santa Fe Depot	235775	207316	Increase	443091	168104	Decrease	274987
2	San Diego–Old Town Transportation Center	113163	140454	Increase	253617	101689	Decrease	151928
3	Solana Beach	94621	81468	Increase	176089	86720	Decrease	89369
4	San Juan Capistrano	71260	46266	Increase	117526	42531	Decrease	74995
5	Anaheim–Regional Transportation Intermodal Center	71651	87145	Increase	158796	29565	Decrease	129231
6	Santa Ana	50733	45121	Increase	95854	23241	Decrease	72613
7	Fullerton	86618	85512	Increase	172130	11508	Decrease	160622
8	San Clemente Pier	7563	6544	Increase	14107	8997	Decrease	5110
9	Macomb–Thomas C. Carper Station	25707	15968	Increase	41675	4476	Decrease	37199
10	Kewanee	6588	5485	Increase	12073	2587	Decrease	9486
11	Mendota	9002	5890	Increase	14892	2245	Decrease	12647
12	Quincy	14450	10590	Increase	25040	2112	Decrease	22928
13	Princeton	12441	9446	Increase	21887	1969	Decrease	19918
14	Pittsburgh–Union Station	71015	46951	Increase	117966	1882	Decrease	116084
15	Essex Junction-Burlington	4091	13321	Increase	17412	1411	Decrease	16001

Below are our graphs from Tableau.



Our Findings

In our findings, out of 525 stations analyzed, 51 experienced a decline in ridership from FY22 to FY23, despite showing growth between FY21 and FY22. This pattern points to a post-COVID adjustment, where the initial surge in ridership after pandemic-related disruptions has leveled off as commuting habits settle. For instance, key stations like San Diego–Santa Fe Depot and San Diego–Old Town Transportation Center experienced notable drops in ridership. These shifts could reflect changes in commuter behavior, operational hurdles, or evolving travel habits in these regions.

Our Recommendations:

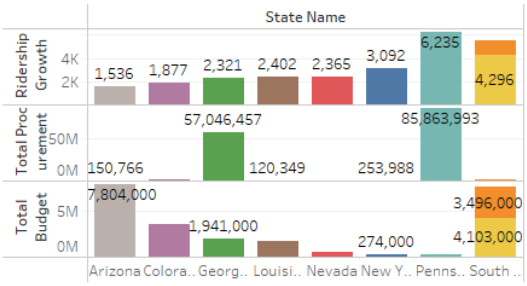
Given our analysis, we would recommend the following:

1. Analyze delays to distinguish between environmental (weather, infrastructure) and human-related (crew shortages, passenger behavior) factors using data-driven insights.

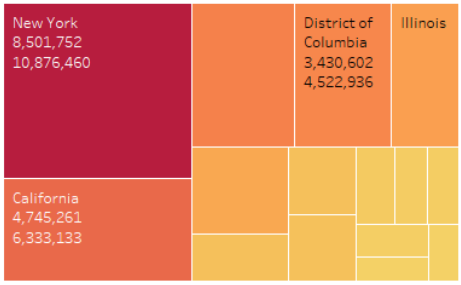
2. Upgrade infrastructure, improve communication, train staff, and collaborate with weather services to proactively reduce delays and enhance customer satisfaction.
3. Incorporate buffer times, use dynamic real-time adjustments, and align train schedules with connecting services to minimize cascading delays and improve reliability.

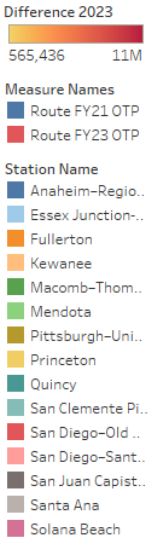
DASHBOARD

Which stations have the lowest and highest ridership growth, and how does this compare to the stations total procurement and total budget allocations?

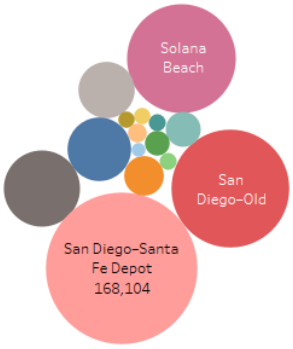


What are the states where ridership exceeded rewards participation in 2023, and how do the differences between ridership and rewards compare across 2022 and 2023?





Which stations saw a drop in ridership from 2022 to 2023, how did their ridership change from 2021 to 2022?



Which are the bottom fifteen routes with the lowest on-time performance (OTP), and what factors contribute to their low performance?