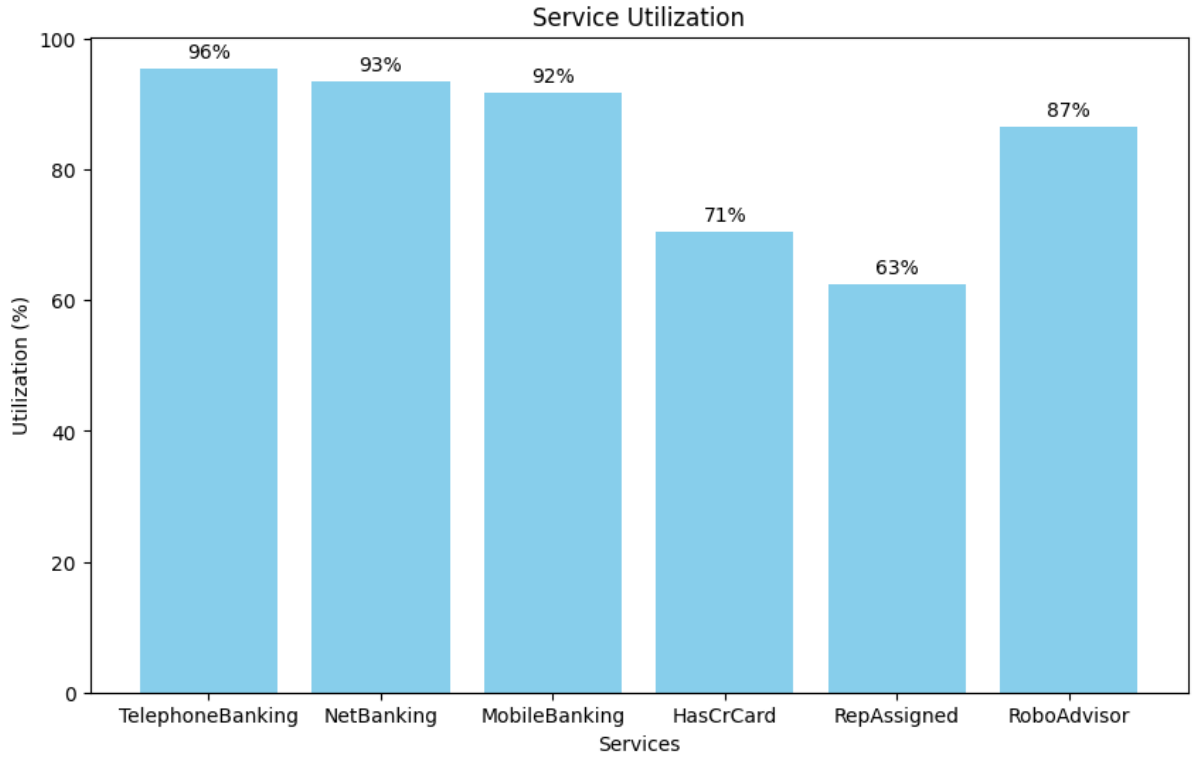
The Bank Co Project

Peter Sarvari

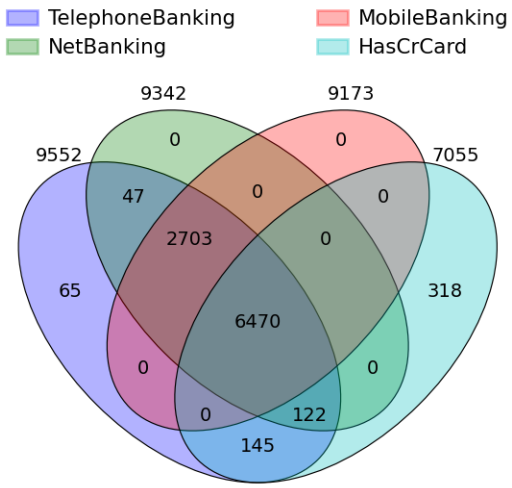
Overview: The Bank Co Dataset contains information about 10,000 unique customers including their personal data and their usage of services provided by the bank. Personal data includes address information, age, current account balance, estimated salary, credit score, tenure with the bank, number of contacts with the bank within the last 12 months and whether the customer closed their account or not. Service usage data include information whether the customer uses certain core services (yes or no): telephone banking, net banking, mobile banking, credit card and whether the customer has a human representative, or a robotic advisor assigned (or neither or both). Based on this data, we're asked to (1) give a visual representation of services used by customers, (2) give a visual overview of the types of customers by region, (3) give a visual overview of what might be driving customers to quit and (4) raise anything else The Bank Co should look into.

1. Visual overview of the type of services used by customers

The core services of the bank examined here are Telephone banking, net banking, mobile banking and credit cards. Services may be provided by customer representatives and robotic advisors. Let's start by looking at the utilization of such core services and service provisions by the 10,000 customers (79.63% active and 20.37% exited).



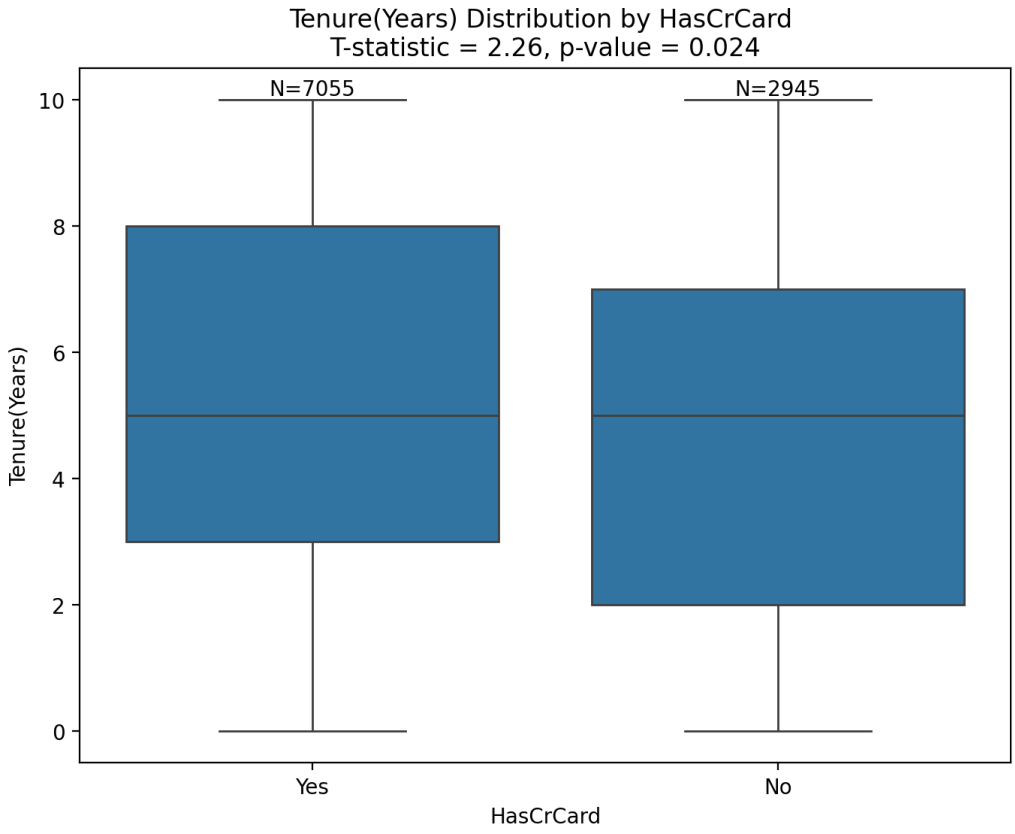
The bar chart shows that credit card is the least utilised core service (71%) and while 87% uses Robotic Advisors, only 63% of customers have a representative assigned. It is important to understand if these core services are useful on their own, or most customers use a combination of them. We can use a 4-group Venn diagram to give a visual answer.



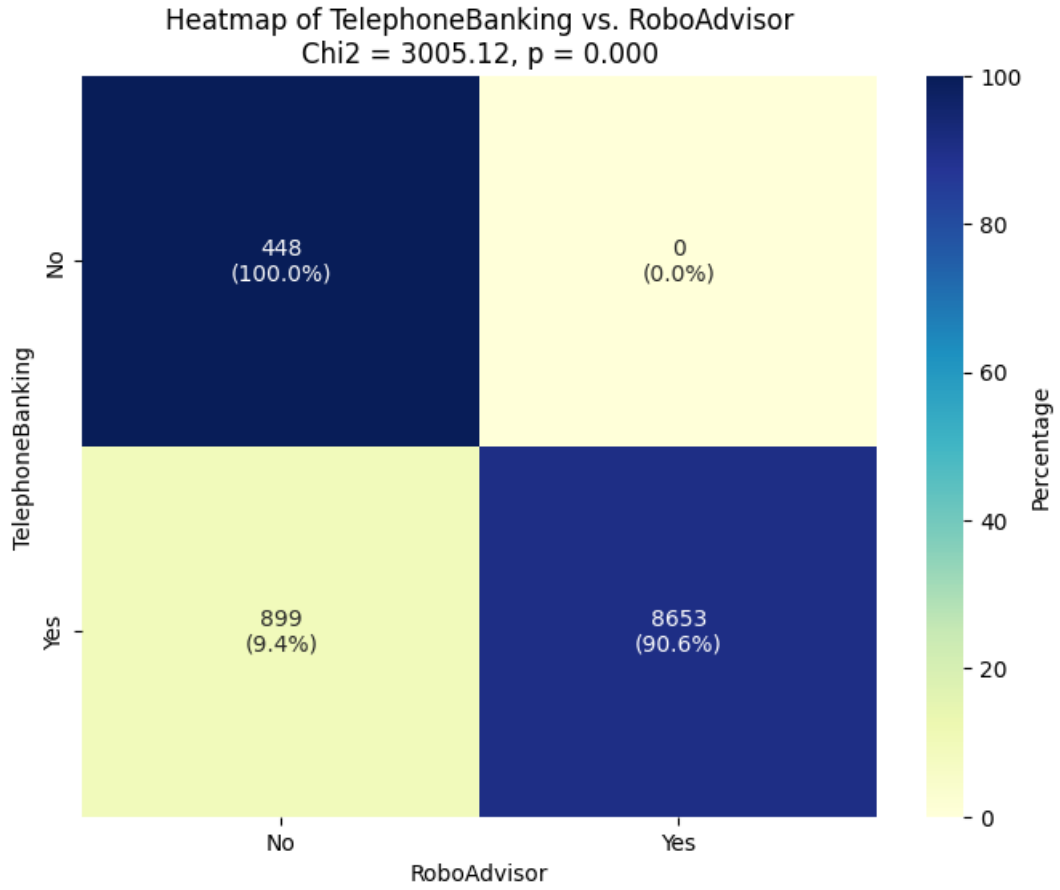
We can observe that only very specific combinations of services are used: 6470 customers use all 4 core services, while 2703 use all except for credit card. 318 uses the credit card only, 145 uses only telephone banking and credit card, 122 uses all except mobile banking, 65 uses only telephone banking, while 47 relies only on telephone banking and net banking. It'd be insightful to understand if customer personal data or service provision (reps/robot) influences the usage of core service. With this goal in mind, we have created boxplots that show the distribution of continuous variables for cases when the core service is used and when it is not used. The following table summarizes the continuous variables that are significantly different (p-value <0.05 using T-statistic) between users and non-users of the given service.

Service	Age	Tenure	CreditScore	Contact_12months	Balance	EstimatedSalary
TelephoneBanking	X			X	X	X
NetBanking	X		X	X	X	X
MobileBanking	X		X	X	X	X
HasCrCard		X				

One example of the 24 produced boxplots is shown below: HasCrCard is statistically significantly associated with tenure meaning that customers who have been with the bank longer, are more likely to have credit cards. This is not necessarily intuitive and is worth looking into and understand if it is necessary for new joiners to be actively asked about a credit card.



Similarly, we have examined which binary variables are significantly associated with the usage of the 4 core services. Interestingly, 3 services, namely TelephoneBanking, NetBanking and MobileBanking were significantly associated with all 4 binary variables examined: gender, exited, RepAssigned and RoboAdvisor, whereas HasCrCard was didn't seem to significantly depend on any of the above 4 binary variables. We created 2-by-2 heatmaps to show the nominal values as well as the percentages in the relevant squares. To calculate the p-values, we used the Chi-Squared test. See an example of the heatmap below.



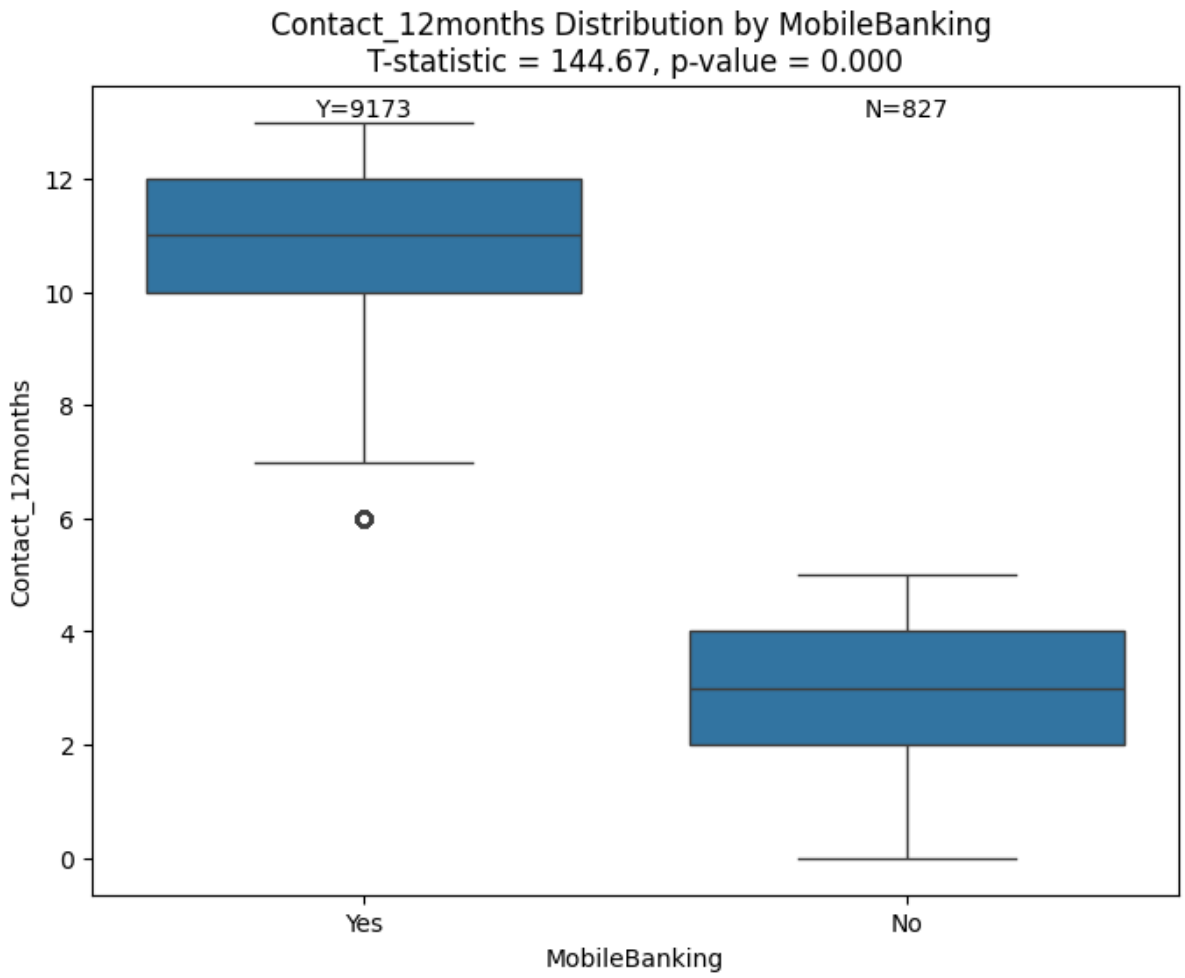
After closely examining the above boxplots and 2-by-2 heatmaps, we realized the following general truth about this dataset:

- Only those who contacted The Bank Co more than 3 times have TelephoneBanking, only those who contacted the bank more than 4 times have NetBanking and only those who contacted the bank more than 5 times have MobileBanking.
- Only those who contacted the bank more than 8 times have a RoboAdvisor
- Only those who contacted The Bank Co more than 10 times have a RepAssigned
- Everyone who is still with the bank (has not exited) has contacted the bank at least 10 times last year

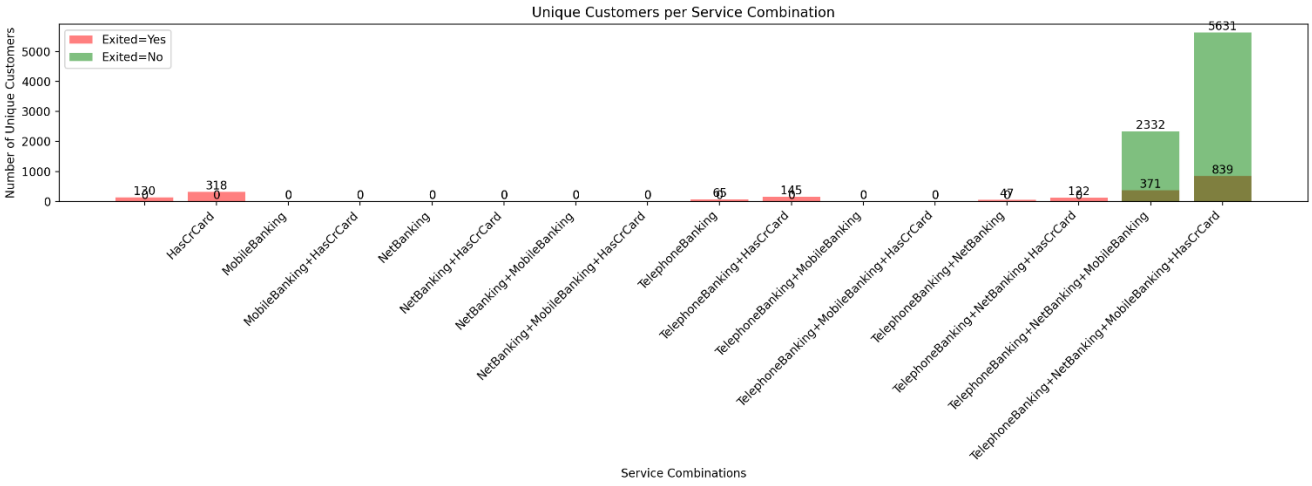
As a result, it follows that

- All customers who have not exited have TelephoneBanking, NetBanking, MobileBanking and RoboAdvisor but may or may not have a RepAssigned

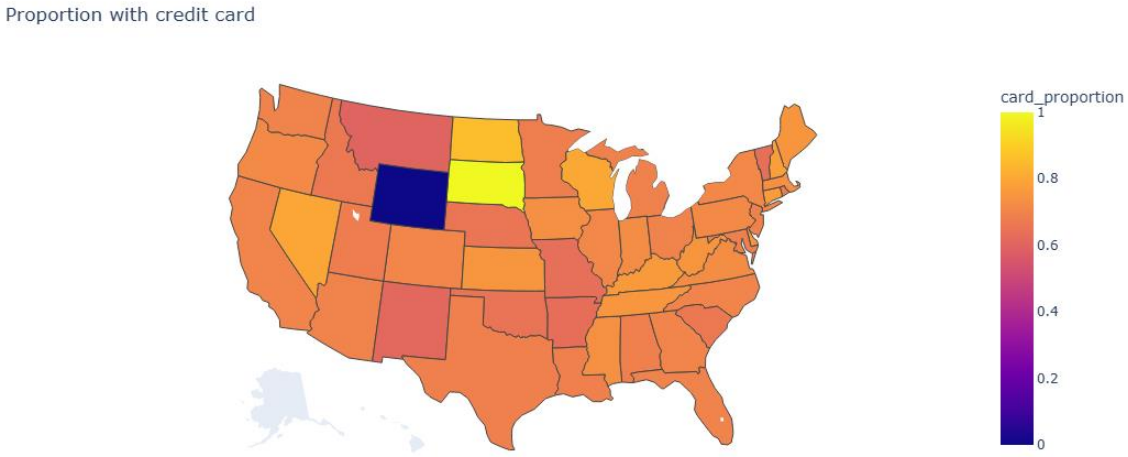
To conclude these findings, boxplots like the one below helped us.



The following plot also confirms the findings and shows a more detailed visualization of the previous Venn diagram broken down by Exited and not Exited users.



While we found many relationships between the variables and the Tele/Net/Mobile banking services, factors determining credit card ownership are less straightforward. Out of all variables examined, only tenure was significantly associated with credit card ownership. The only variable we haven't examined yet (but will soon in the next section) is location. Below we will show the probability to hold a credit card if you're a Bank Co customer in each state.

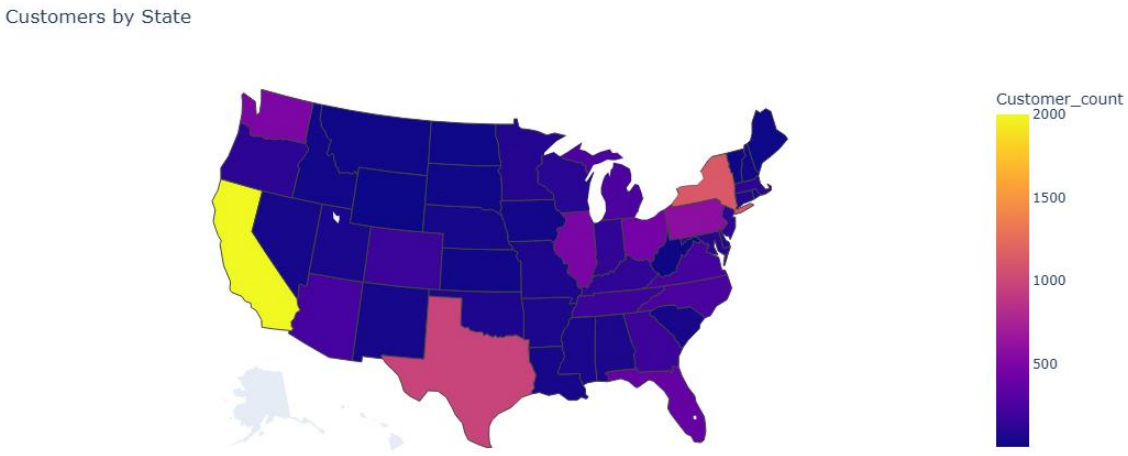


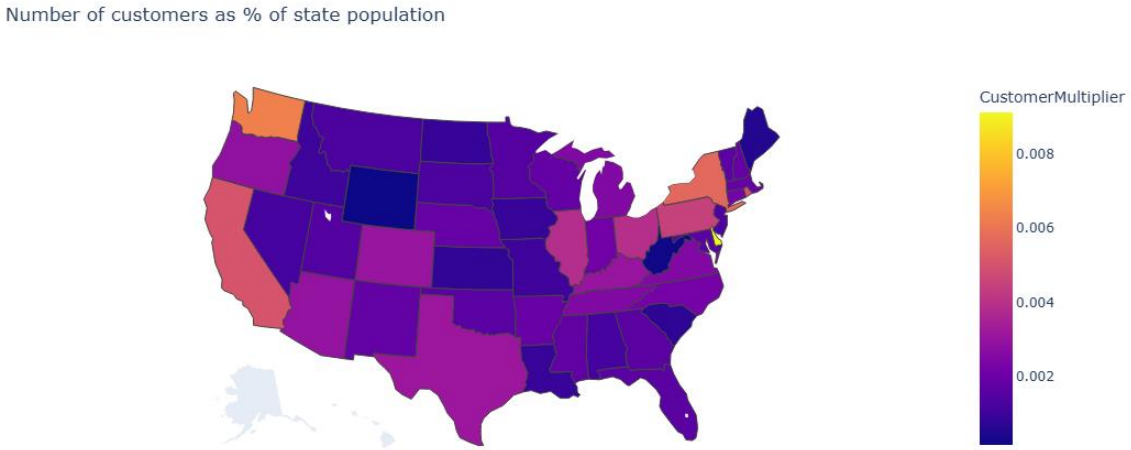
Proportion of customers with credit cards fall within 0.6 and 0.8 in all but 3 states: Wyoming (zero but we only have one customer), South Dakota (all 12 customers have credit cards – might warrant digging deeper to understand why this is the case, but maybe sales simply did well) and North Dakota (0.86: 5 out of 7 customers have credit cards). The probability to hold a credit card seems to be relatively uniform across states with a reasonable number of customers.

2. Visual overview of the types of customers by region

First, we will show the number of customers and the customers' average age, balance and salary by State on the map. Then, we obtain average age (*Wikipedia, 2022*), average salaries (*Knitpeople, 2024*) and total population (*Wikipedia, 2020*) in the various US states, which allows us to calculate the average age multiplier (showing the ratio of average Bank Co customer age to the average age in that state), average salary multiplier (the same but for salaries) as well as the number of customers as percentage of the state population. While the maps pasted here are color-coded, they are interactive via the notebook pasted in the Appendix, meaning that the exact number or ratio is shown by hovering over the state.

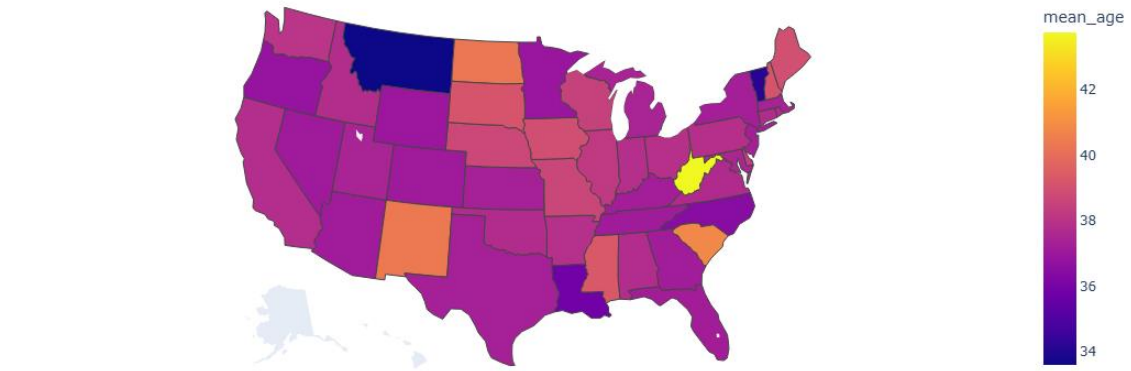
a. Customers



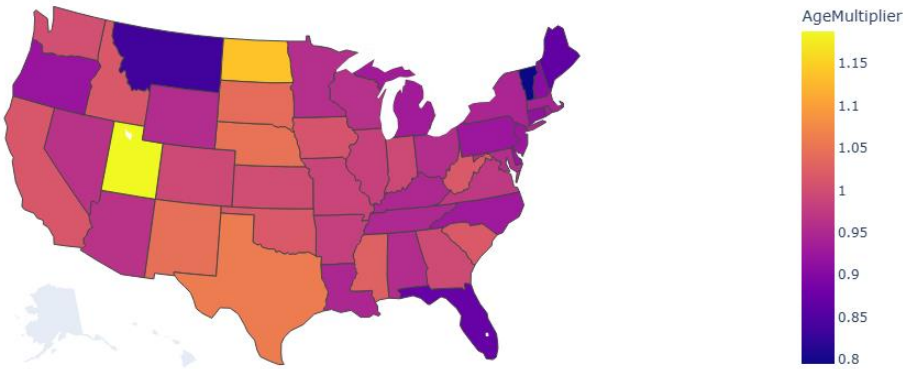


These plots tell us that while California has the most customers (2002), The Bank Co captures the highest ratio of the state population in Delaware (96 customers corresponding to 0.009% or 0.9bp while in California the same metric is only 0.5bp). It might be worth investigating why the conversion is notably higher in Delaware than the rest of the states: are people there simply more suited for The Bank Co products or is it the marketing/sales effort in Delaware that makes it succeed? The first heatmap visually tells that there are only a handful of customers in Central and Northern US except for Colorado and the Great Lakes area. We've also noticed that The Bank Co has only a single customer in Wyoming. We're wondering how they acquired that customer, Mr. Angelo (maybe they recently moved from another state using net banking) and whether The Bank Co is actively looking to expand in Wyoming.

b. Age



Average Age Multiplier

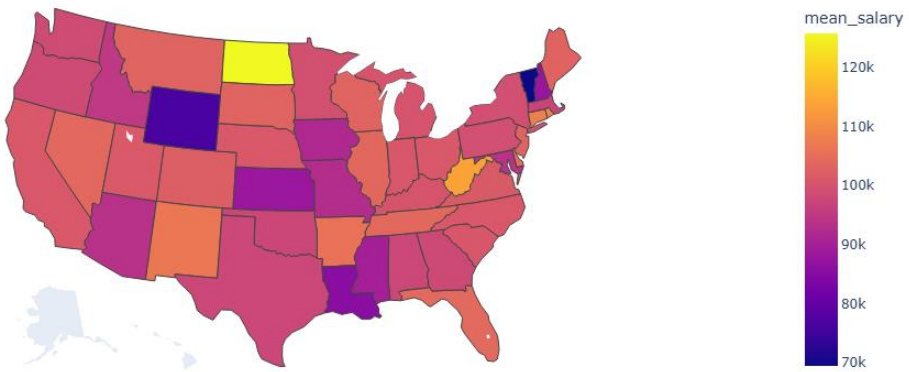


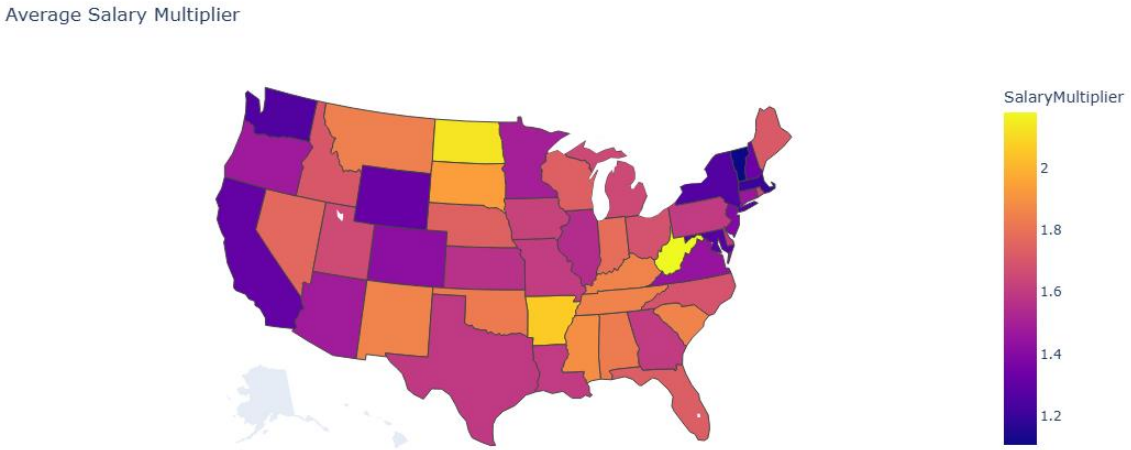
We observe that there are states where younger than average customers use The Bank Co (e.g. Vermont and Montana), while older than average customers use the services in other states (e.g. Utah and North Dakota). Age distribution also varies a lot across the states (33.6 in Montana vs 43.75 in West Virginia) but the lack of customers in these states (15 in Montana and 4 in West Virginia) could skew the averages and make these findings not conclusive. Regardless, it'd be interesting to understand if The Bank Co appeals to different age groups in different states, and if so, then why.

c. Salary

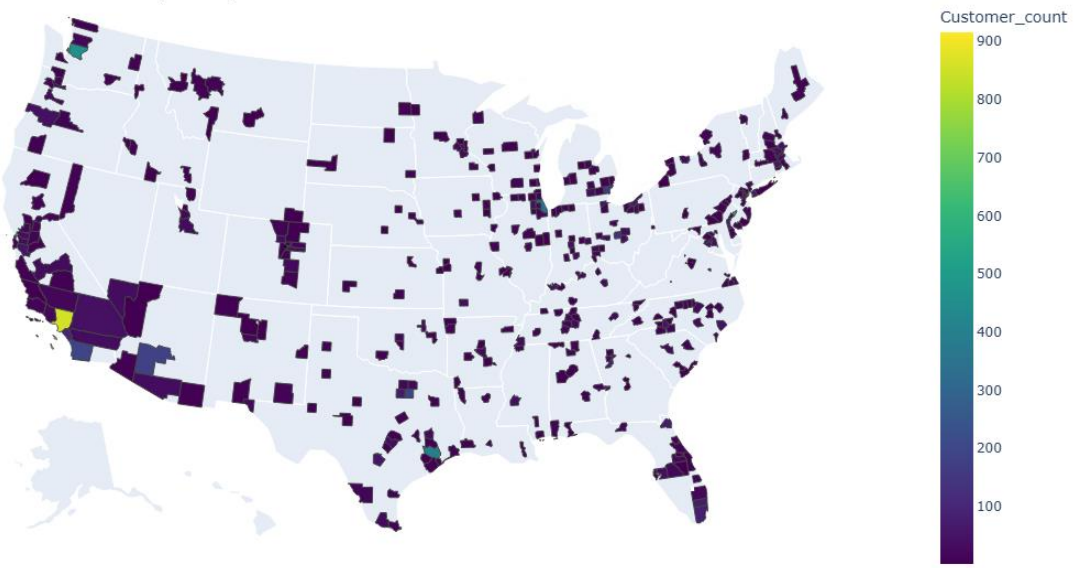
Regarding salary, a similar trend is observed on the maps below. Both nominal salary and salary range varies widely across state, potentially exacerbated by the lack of customer data in some states. It is, however, interesting to observe that the average customer's salary is at least 1.1 times the average salary in that state, suggesting that The Bank Co is primarily focusing on wealthier customers.

Average Salary



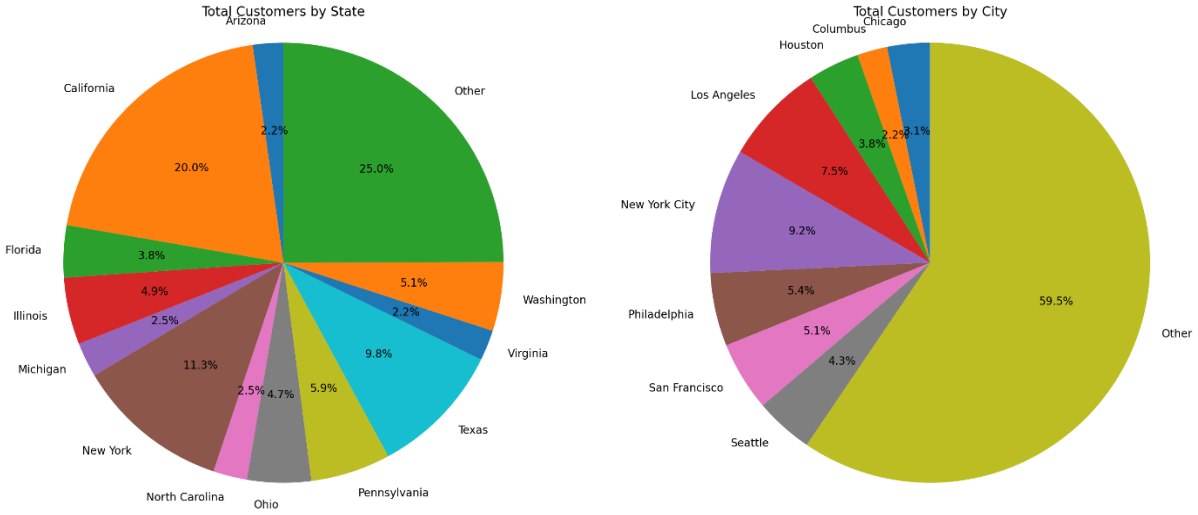


Finally, we wanted to show the customer count broken down by county level.



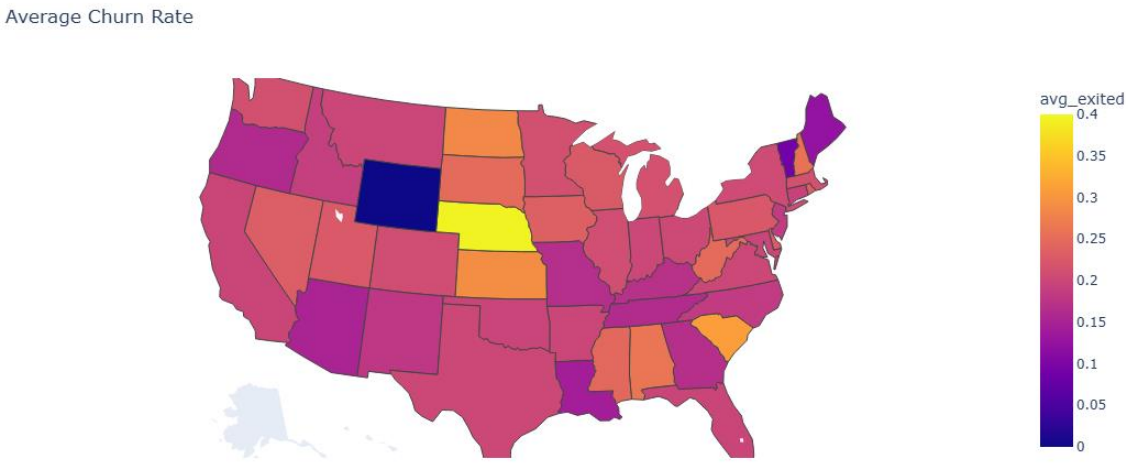
This clearly shows that the customers are not uniformly distributed within the states and there is potential for the bank to acquire new customers from "no-customer" counties nearby "hotspots" and understand why it wasn't able to do so thus far. Not surprisingly, the top counties in terms of customer numbers are New York and Los Angeles corresponding to the two largest cities in the US (*Wikipedia, 2019*). However, the county that the 3rd largest city, Chicago is situated in (Cook County) only comes 7th preceded by Philadelphia, San Francisco, King and Harris counties.

The total customers by state and by city are summarized in the following pie charts.

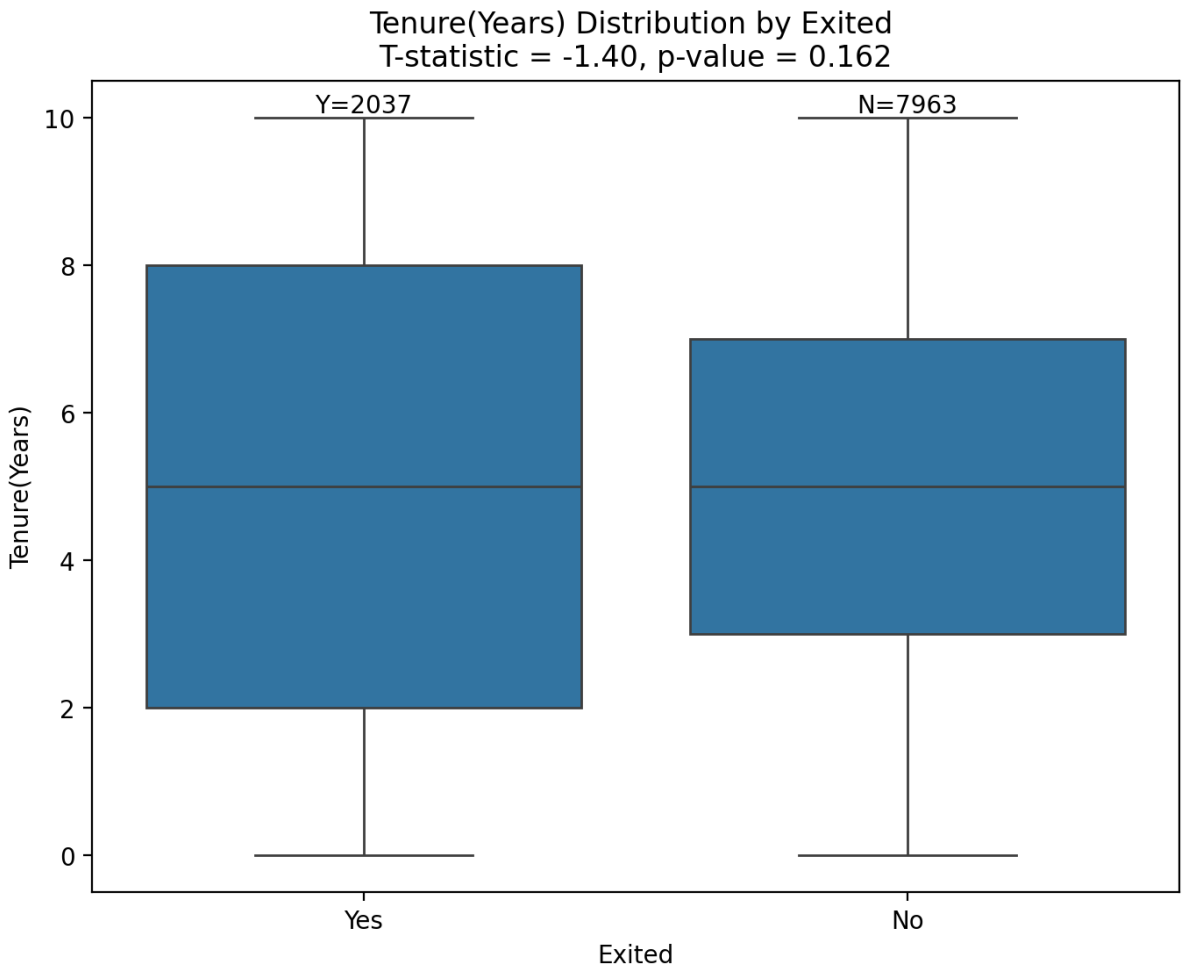


3. Visual overview of drivers of customer churn

The average churn rate across the states varies substantially ranging from 0 (in Wyoming, albeit only one customer) to 0.39 in Nebraska (only 38 customers).

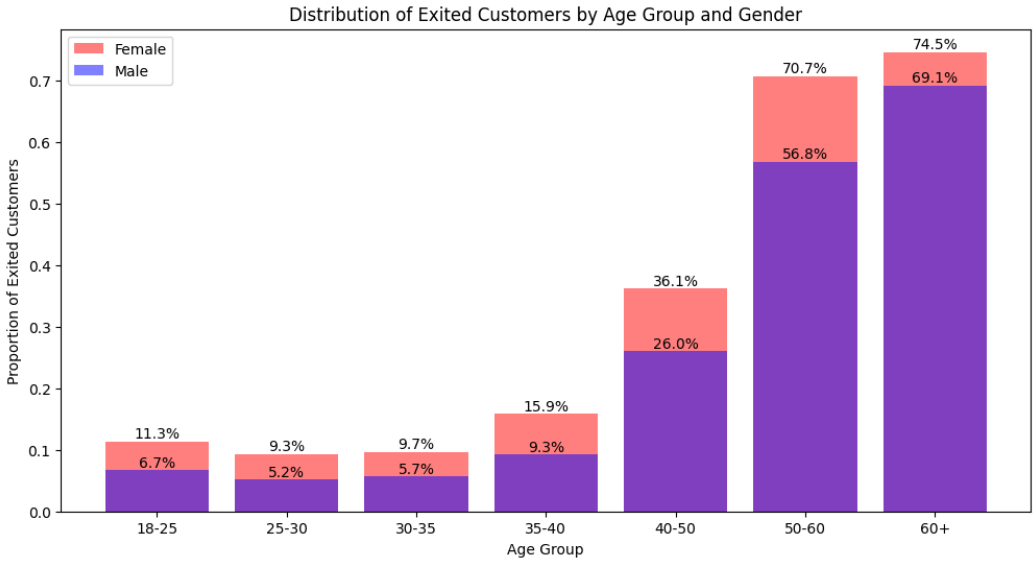


The findings that we discussed in section 1 relate the Tele/Net/MobileBanking services as well as service provisions to Exited status via the Contact_12months variable. In essence, because all customers who have not exited, contacted the bank at least 10 times, they have Tele/Net/MobileBanking services enabled as well as a robotic advisor and they may have representative assigned as well. So, of course, all these variables will be statistically significantly different between customers who exited and those who have not. Additionally, gender, account balance, age and credit score are significantly associated with Exited status: older age (p-value < 0.001), lower credit score (p-value = 0.007), higher balance (p-value <0.001), being female (p-value <0.001) and of course, lower 12-month contact numbers (p-value < 0.001) imply more likely to exit. We found it quite interesting that tenure doesn't seem to be significantly affecting exit status (see below).



Normally, we'd expect that with time, people become more loyal to the bank and are less likely to leave. Alternatively, banks may apply new joiner discounts (no fee for the first year, cashback, higher savings account rate, etc) in which case, we may observe the opposite: the longer the tenure, the more likely that customers quit. In case of The Bank Co, however, we don't particularly observe any of these characteristics.

We also examined the distribution of quitters (exited customers) by age, grouped by gender:

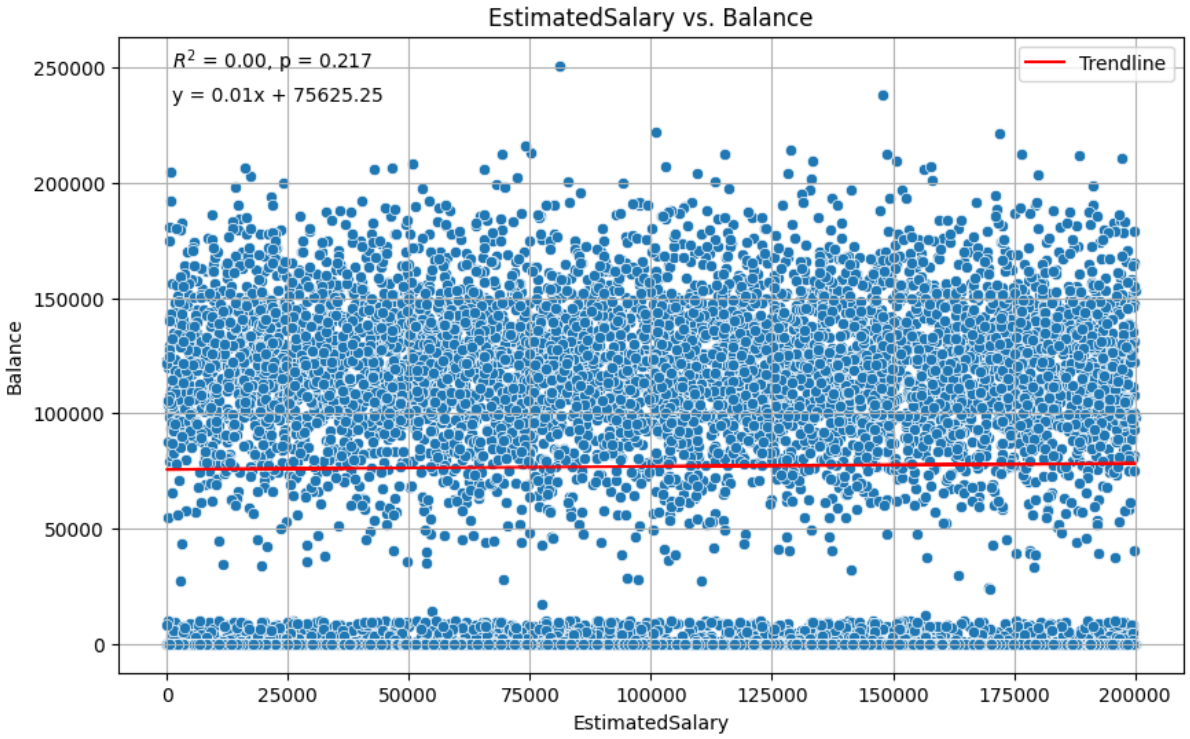


It is clearly visible that for every age group, females are more likely to exit and that the exit probability grows by age. As the exit probability effectively doubles from the 40-50 group to the 50-60 group, The Bank Co should really prioritise finding strategies to reduce the increase in churn in older groups.

4. Further findings / discussion

Apart from services, customer locations and churn, we have taken a look at service provisions. We were surprised to find in section 1, that RoboAdvisor and RepAssigned seem to be completely determined by the number of contacts the customer had. Why does to bank add these service provisions to frequently contacting customers only? Do customers with a representative still need the robo advisor? If we were to design service provisions, as consultants, we'd help the poorer clients with lower credit scores with the robo advisor and would provide exclusive service to high-net-worth clients by assigning customer reps for them.

In addition, we looked at the relationships between salary and balance as well as age and balance. Remarkably, no relationships were found between these variables as the below graphs suggest.



These may suggest that the customers of The Bank Co are not too keen on savings – even if they earn more, their balance is not more likely to increase (they might spend more) and even as they earn money longer, their balance doesn't increase significantly (they have more time to spend it). While these are preliminary findings, it'd be interesting to explore these relationships more as this should influence the marketing strategy of The Bank Co: for example, if customers are not interested in saving, cashbacks may be a better incentive than interest earned on savings accounts.

We also investigated the relationship between credit score and estimated salaries (p-value: 0.89) as well as credit score and balances (p-value: 0.55), however, to our surprise, no significant relationships were found.

Appendix

Python Jupyter Notebook is available from: https://colab.research.google.com/drive/10E_Ea92-e2LF0BDXCXasG_QazaU8yFm0?usp=sharing

References

Knitpeople. (2024, June 13). *2024 US Average Salary by State*. Knitpeople.com. https://www.knitpeople.com/blog/average-salary-us

Wikipedia. (2019, August 25). *List of United States cities by population*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

Wikipedia. (2020, April 1). *List of states and territories of the United States by population*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_population

Wikipedia. (2022, April 20). *List of U.S. states and territories by median age*. Wikipedia. https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_median_age