A picture containing text

Description automatically generated

ALY6080 INTEGRATED EXPERIENTIAL LEARNING

TECHNIAL COOKBOOK

Logo

Description automatically generated

Prof. Kristen Drobnis

Submitted by Team 2:

Anar Gochiyev, Anil Singh Chouhan,

Ashish Kumar, ​Pranay Chidura, Wanlu Yao

December 12, 2020

**Introduction**

Data is the thing that every organization uses to build, improve, and achieve the best result in their business. Today, data helps to make decisions, to run business, to increase revenue, to make things easy, and to collaborate. Therefore, every organization needs data about the business, people, contacts, and offices as our sponsor (NextPet) also needs an updated list of contacts. To get updated contact we have used different web scraping techniques to provide our sponsor to help their business grow. Scrapping the data from multiple sources gives us data with many ambiguities to deal with and the structured data help us to provide our sponsor with a dashboard that easily gives insights into our research and data collection.

Our sponsor, NEXT PET LICENSING SYSTEM is the first-ever mobile application software and a start-up founded in 2015, which deals with pet licensing services. They assist municipalities with building more advantageous, more secure, and more astute pet networks by expanding pet authorizing compliance and animal control programs. NextPet is the most complete pet administration stage with elite other exclusive descriptions such as collaboration with veterinarians for rabies shot upload, work-flow automation, multi-language support, google maps, and Stripe APIs integrations. NextPet is elite customizable based on every network need. It provides the cloud-based services to help the user get connect from anywhere, to anyone and get their pet licensed on fingertips.

Next Pet is a newly built start-up that deals with animal control services and helps to improve the animal control system. It aims to collaborate with users with public officials on a fingertip with the use of their cloud-based system. But with an outdated technology and system services of company which deals with pet care are facing problems. On average 40% of dogs are not licensed resulting $720 million loss annually. Due to the unavailability of data municipalities are facing problems in collaborating with pet owners. Therefore, our sponsor aims to improve pet licensing, focusing on the mobile app market and to increases products and services for the animal care system. It aims to improving the connection between the user and public user and focusing on improving the time-consuming process of licensing a pet. Hence, to support our sponsor we are providing them with an updated list of contacts of public officials working in different municipalities throughout the USA. Also, we are providing them recommendations of using of the automated process of collecting information which they need and at last giving an interactive dashboard to help grow their business.

**Data Collection Cookbook**

*Data Collection Method I*

The first data collection tool we found is named “Agenty”. Agenty is a SaaS platform with easy-to-use automated agents in cloud for machine intelligence. With its API ready agents, it allows business and professionals to extract, clean, validate, translate, OCR, sentiment analysis, text classifier, change tracking the big data for businesses across Retail, Healthcare, Machine Learning, Artificial Intelligence and more. For the data collection purpose, we used the scraping agent. It allows users to create personalized web scraper online and run it on Agenty web scraping software on cloud. Using the scraping agent, extract data from any website or scrape the data from thousands of websites in minutes and converts unstructured data into structured spreadsheets and machine-readable data with no coding required.

Step 1:

The first step is to create an account and install extension in chrome from Google Chrome store.

Graphical user interface, text, application

Description automatically generated

Step 2:

Once the extension is installed, go to the web page you want to scrape the data from. Then, launch the extension and it will display a panel in right side as in this screenshot.

Graphical user interface, application

Description automatically generated

Step 3:

When the panel is visible, click on scraping agent under “create new”.

Graphical user interface, application

Description automatically generated

Step 4:

Next, you can give a name to your field and then click on the yellow asterisk button to enable the selector. The automatic CSS selectors will generate the HTML element automatically when to select the information you want to scrape. For example, when I click and select “contact information” in this site, the selector will highlight and select all matching element with same HTML structure.

Graphical user interface, application

Description automatically generated

Step 5:

The extension will highlight the matching result and will also show you the result preview under the field. Therefore, you can once click on the information you don’t want and verify the result under preview. Once the highlight changes the color from yellow to red, it means the information has been deselected successfully. After the selection is done and you are satisfied with the result, click on the “Accept” button to save the field in your scraping agent configuration.

Graphical user interface, application

Description automatically generatedMoreover, you will have the option the select the information type. The scraping agent provides three extract types, and they are HTML, TEXT, and ATTR.

To scrape URL hyperlinks from websites, you need to select the ATTR option in type and enter href in the name of attribute box, to tell Agenty that you want to extract the value of href in output instead the plain text or HTML. To scrape images from websites, you will also need to select ATTR as the option and enter src in the attribute text box, to tell Agenty that you want to extract the value of src in output for images scraping.

Step 6:

Once you are done with setting up all the fields in your agent, click on the Save button to save your web scraper into your Agenty account.

Graphical user interface, text, application, chat or text message

Description automatically generated

Step 7:

Then click on the link it provides you, and it will be directed you to the home page where all scraped information is saved.

Graphical user interface, application

Description automatically generated

Step 8:

To download the scraped information, click on “Download” button and select the file type to save the file locally.

Additionally, Agenty also allows users to crawl multiple pages with similar web structures at the same time. All you need to do is to go to the input tab and select Manual URLs and enter the URL list in input and save the input configuration.

Graphical user interface, text, application, email

Description automatically generated

More information can be found through this link: <https://www.agenty.com/products/scraping-agent/>

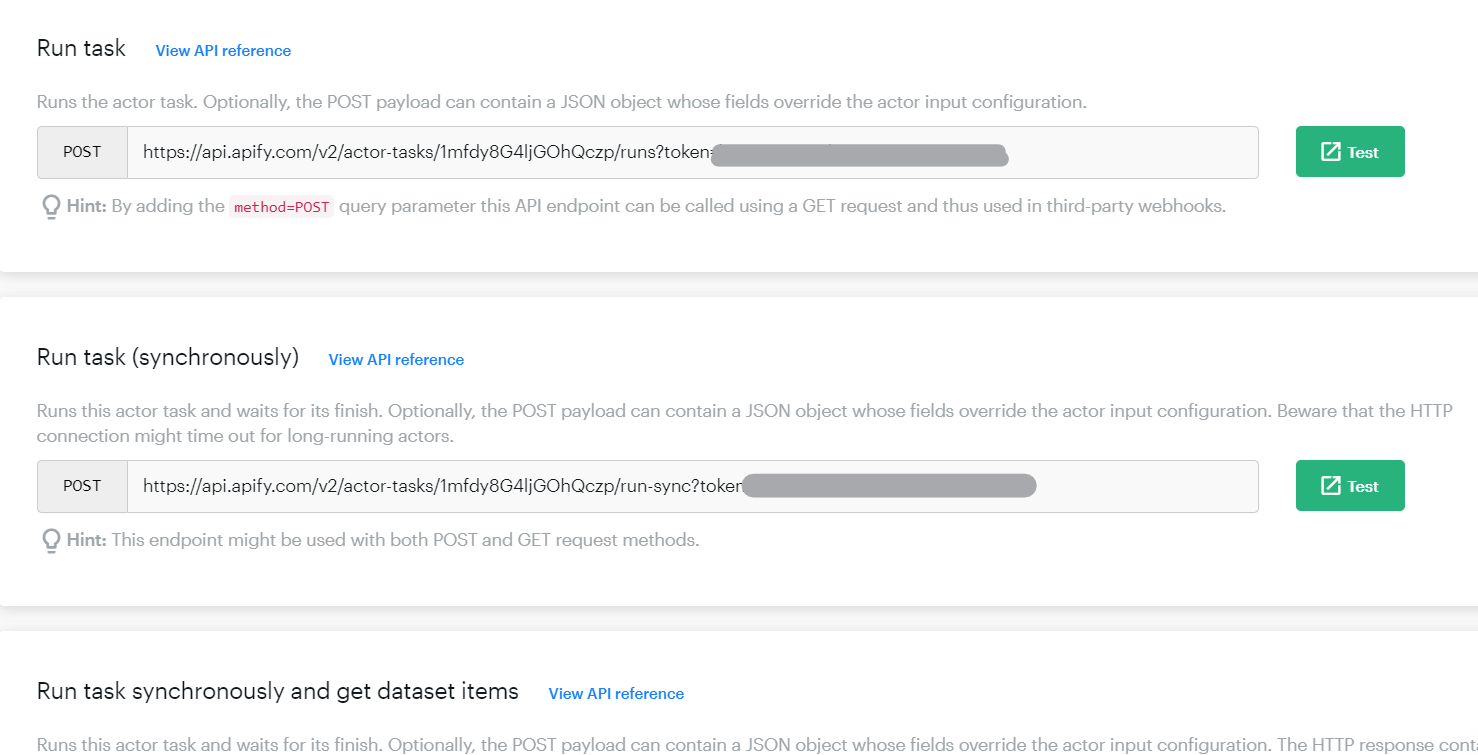
*Data Collection Method II*

Here, we have used a third-party program as web scraping tool for collecting the contact details. To achieve this, we need to follow three steps.

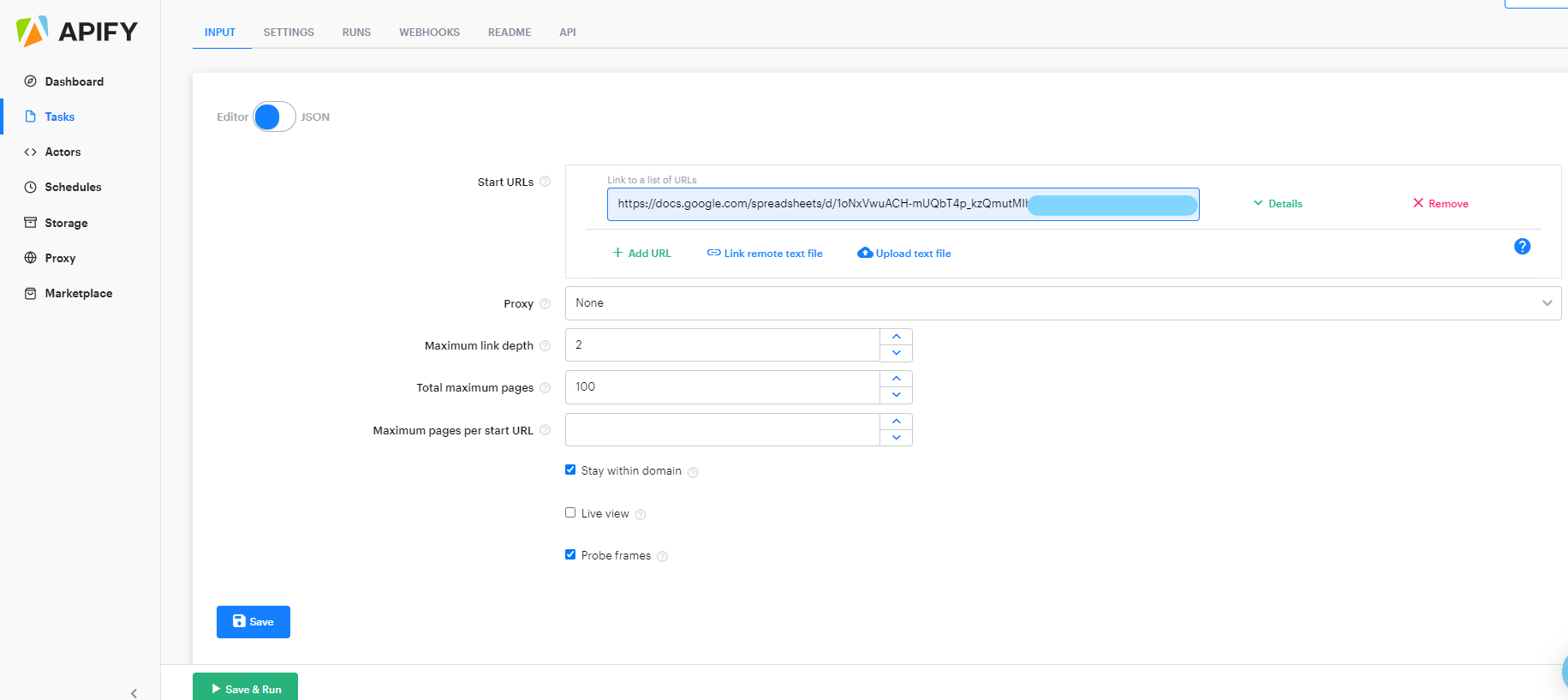
1. Enter URLs in Google Sheet
2. Create a Mini Website and connect through APIs
3. Configure Apify Console

In the first step we need to create a Google sheet and enter the URLs where we want to perform web scraping tools for collecting the data. We can share the document link to specific people who can update the list.

In the second step we need to create a simple mini website such that we can call all API from one place rather than remembering all the APIs.

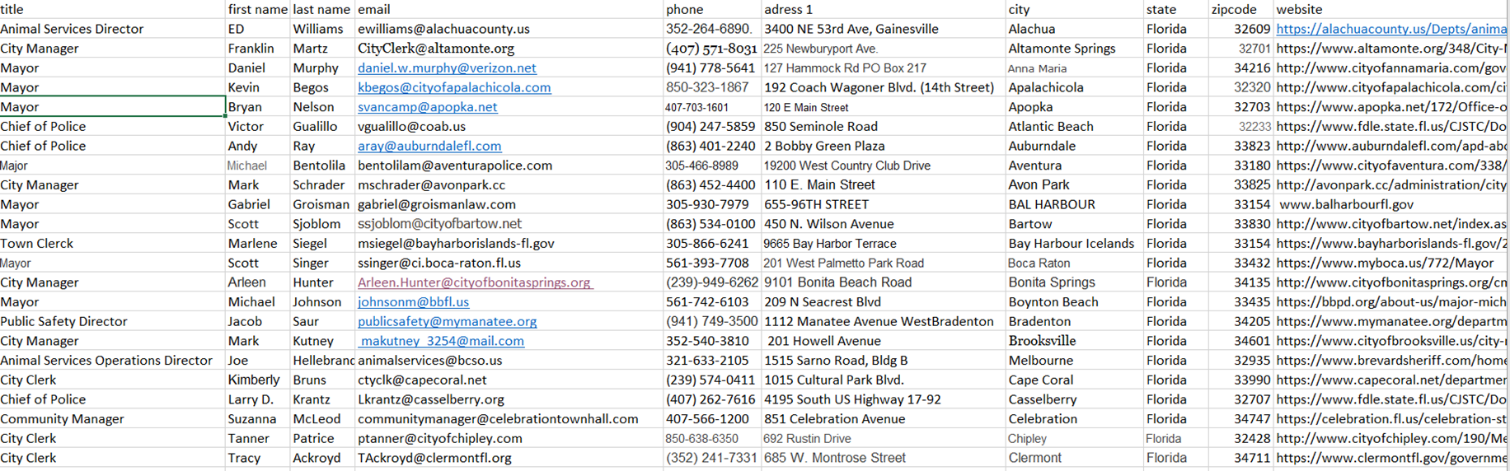


Finally, we need to configure the properties like Link Depth, Max Page Limit to prevent collecting the non-relative data. The Link Depth takes care of how many domain needs to be changed at max while data collection whereas Max Page Limit allows collecting data from certain amount of pages in one URL.

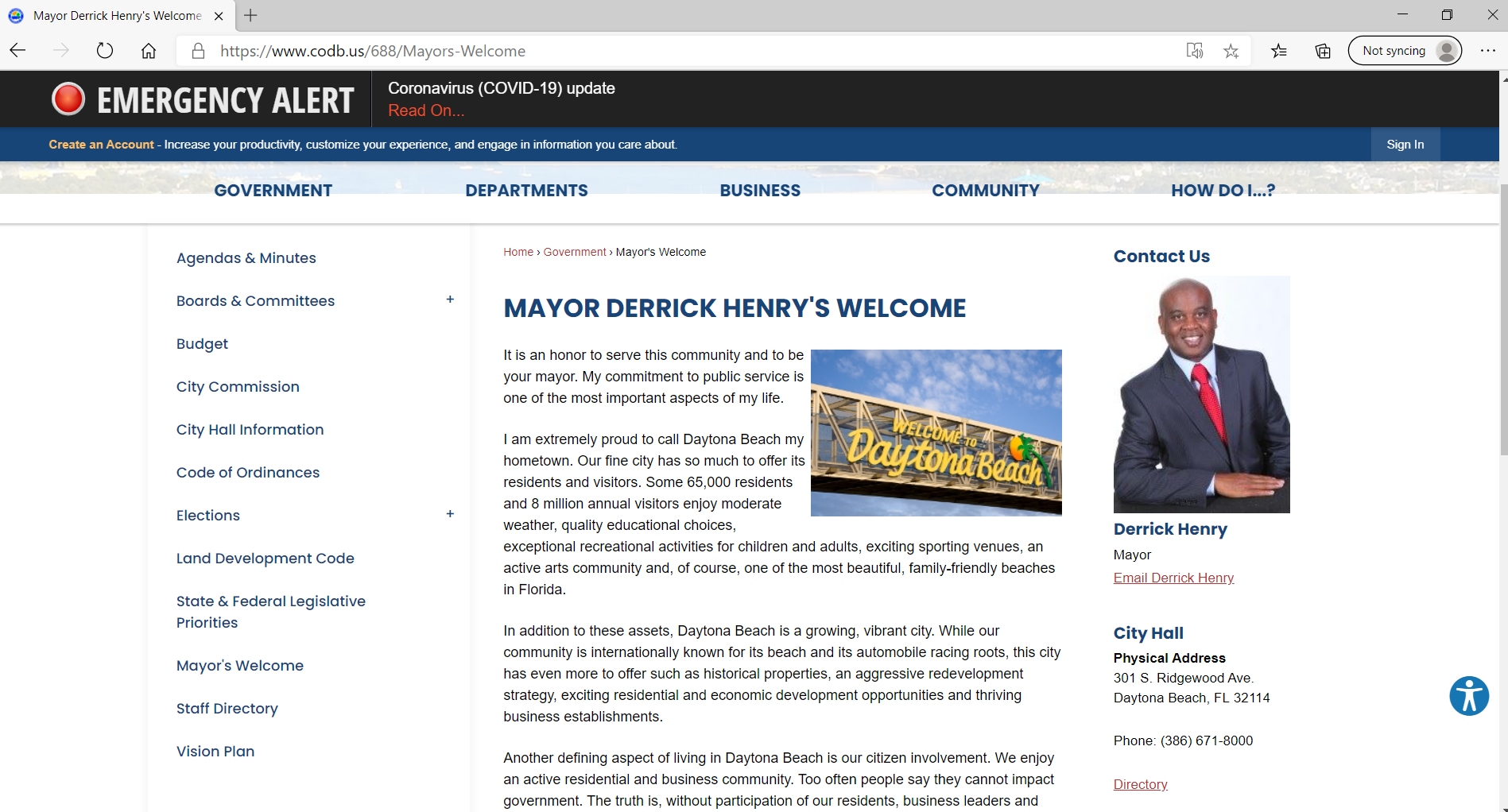


*Data Collection Method III*

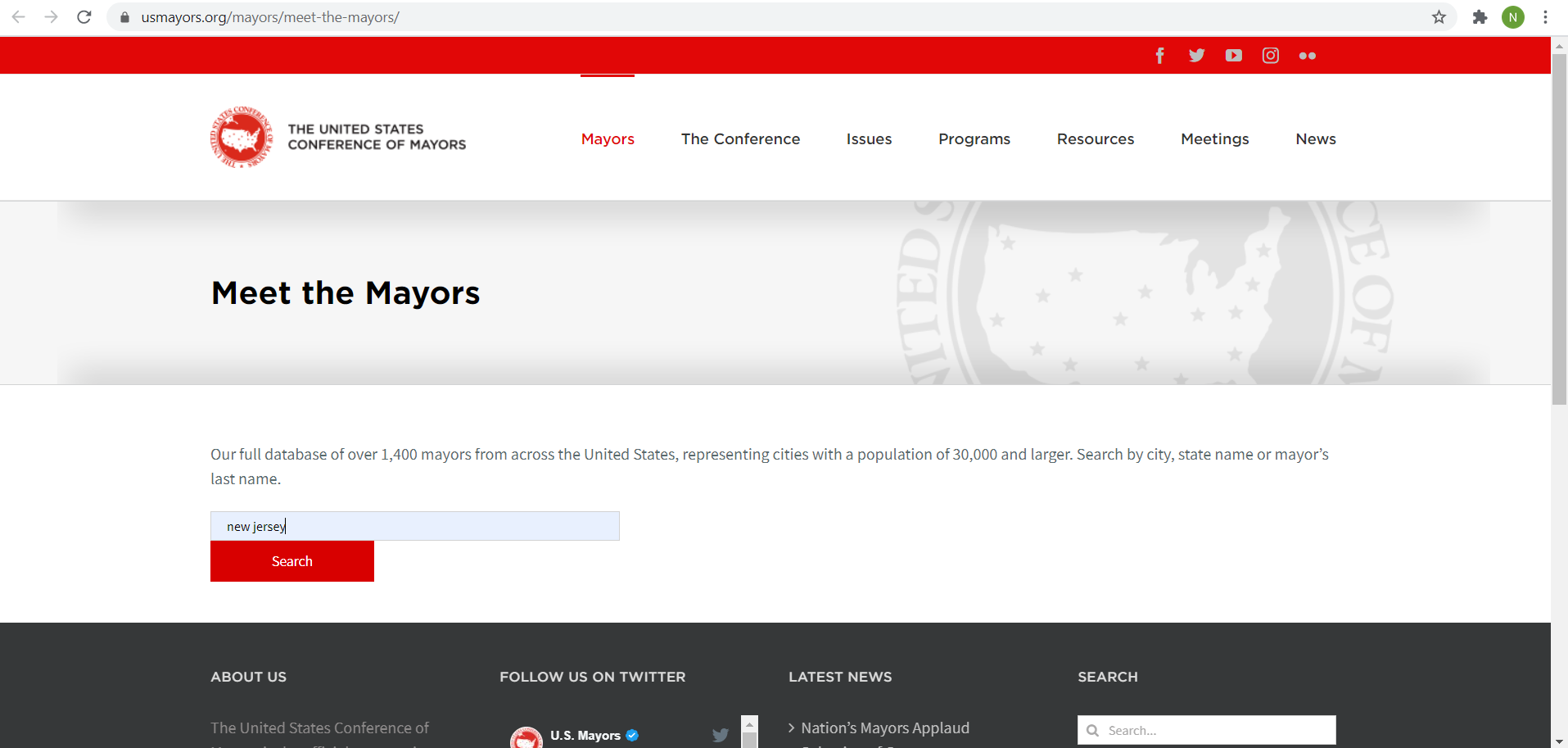
Although in general we collected most of the data by using web scraper tools, we did some manual work too. The reason is to see whether there is any position in municipalities who specifically deal with pet licensing. Based on analysis of more than 100 official websites of municipalities and cities, we come to conclusion that there are not contact details of specific position in every country who deal with pet licensing or animal control.



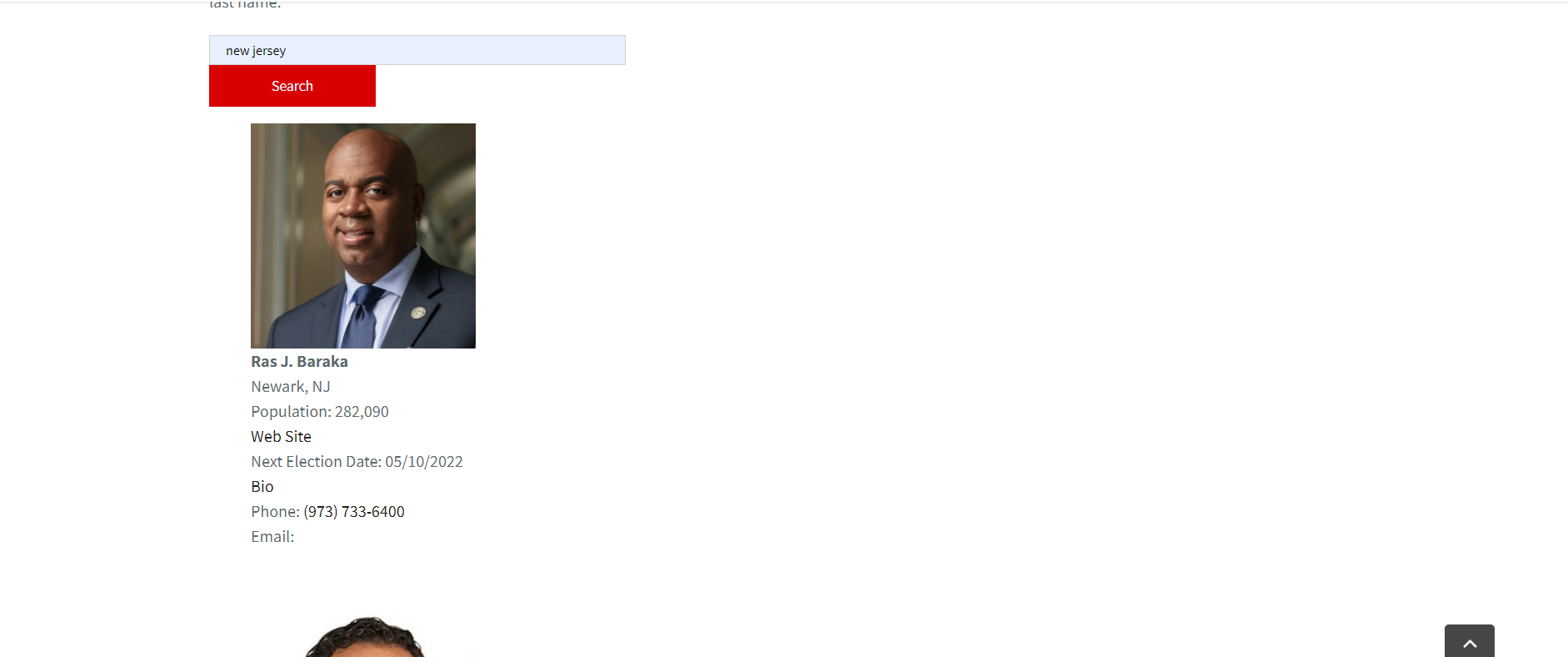
All manual data is collected from official websites of municipalities, which can be considered as reliable source. Here is an example of manual data collection, where we got all the contact information from official websites of municipalities and put them together in Excel file.



Additionally, we used the website of US Conference of Mayors, where updated contact information of more than 1400 mayors from different cities, which have more than 30000 population, are indicated. We collected approximately 100 data from this website but in the recommendation part, we will advise Next Pet to use this website as a database for getting contact information of at least 1400 mayors across USA.



After you entered the website of USA Conference of Mayors, there is search cursor, where you can write a name of state, city or municipality, as a result you will get all updated contact information. Following is the example of search result for New Jersey.



**Dashboard Cookbook**

One of the deliverables for sponsors is an interactive dashboard. In this section, we will walk you through the steps you might perform to create dashboards in Tableau. In order to create a dashboard, you need to build the visuals that you want to add to your dashboard first. For our dashboard, we created six visuals in total.

Step 1: Connect to your Data Resource

When you open Tableau Desktop, the start page shows you the connectors available in the left Connect pane where you can connect Tableau to a file or a server. Depending on the type of your file/ server, click on the option that fit your needs and then connect them.

Graphical user interface, application

Description automatically generated

Step 2: Add More Data Source

Since we want to compare the difference between the sponsor’s data and our final data, we need to connect both data source to Tableau. In the last step, you have connected one data source already, and in the second step, you can add more data source by click on the New Data Source button and select your file/ server path.

Graphical user interface, text, application, chat or text message

Description automatically generated

Step 3: Build Visuals

The second step is creating worksheets and start building visuals. You can rename the worksheet by double clicking on the sheet name and type the new name.

For the first two visuals we built, we want to discover the different of the coverage of contact information between sponsor’s data and our final data, so we built two maps.

1. Before creating the visuals, we need to handle the missing values in our dataset first. Drag State into the Filters pane > unclick on Null in the Default Number Format dialog box.

Graphical user interface, application

Description automatically generated

1. In the Data pane, double-click State, and you will see the State field is added to Detail on the Marks card, and Latitude (generated) and Longitude (generated) are added to the Columns and Rows shelves.

Map

Description automatically generated

1. The default map type in Tableau is often a point map. However, it's easier to see the coverage in polygons. On the Marks card, click the Mark Type drop-down and select Filled Map.

A picture containing diagram

Description automatically generated

1. Then we can add visual details by adding labels to the view, and all you need to do is just drag State to Label on the Marks card.

A picture containing diagram

Description automatically generated

To build a second map based on another data source, just select the new data you want under Data pane, and repeat the same steps mentioned.

For the third and fourth visual, we use the histograms to show the number of contacts.

1. Convert Email and Phone from dimension to measure.

Graphical user interface, text, application

Description automatically generated

1. Double click on Email (Count (Distinct)) and Phone (Count (Distinct)), and drag CNTD(Phone) from columns to rows.

Graphical user interface, application

Description automatically generated

1. Click on the right corner “show me” and select the horizontal bars

Graphical user interface, application, table, Excel

Description automatically generated

1. Then you switch the position between columns and rows, and you will see the bars become vertical.

Chart, bar chart

Description automatically generated

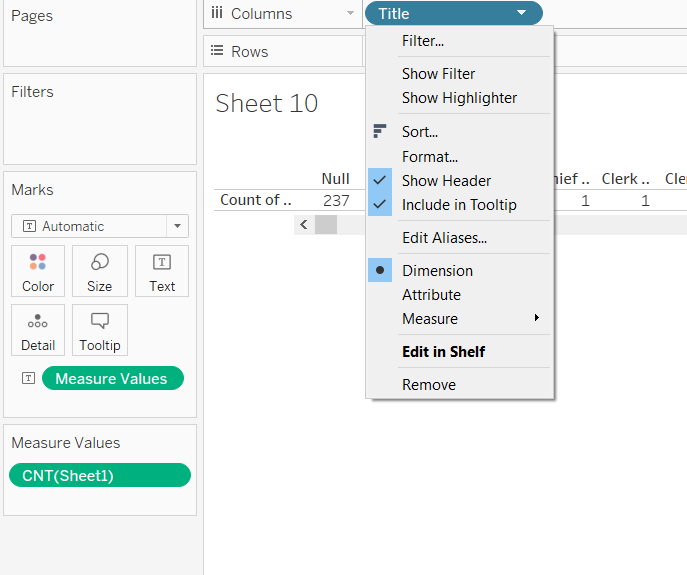
1. Next, drag Measure Values to Label on the Marks card to add details to our graph.

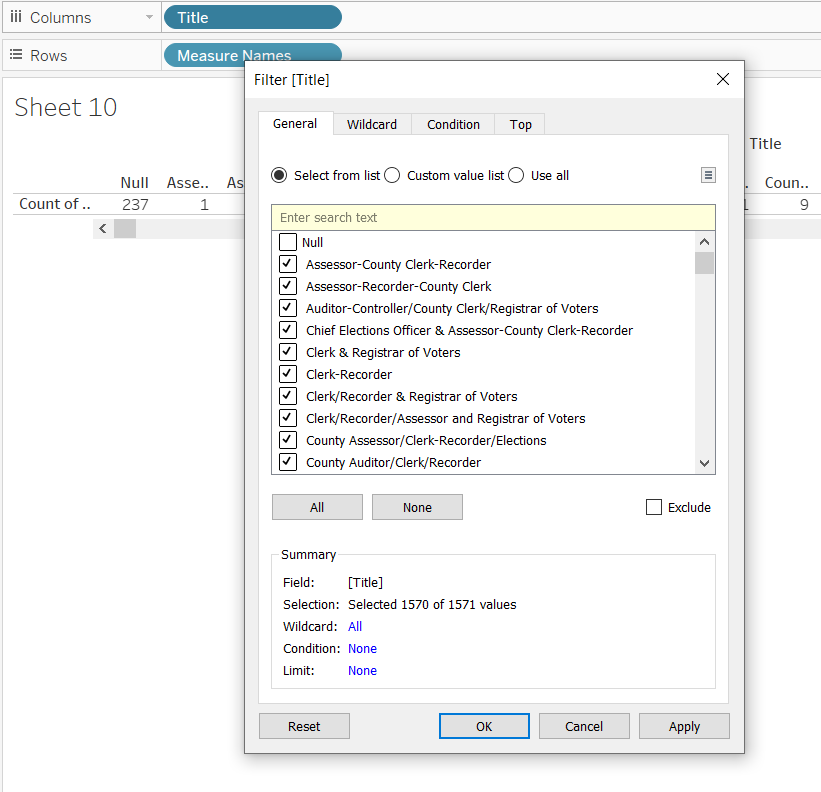
Chart, bar chart

Description automatically generated

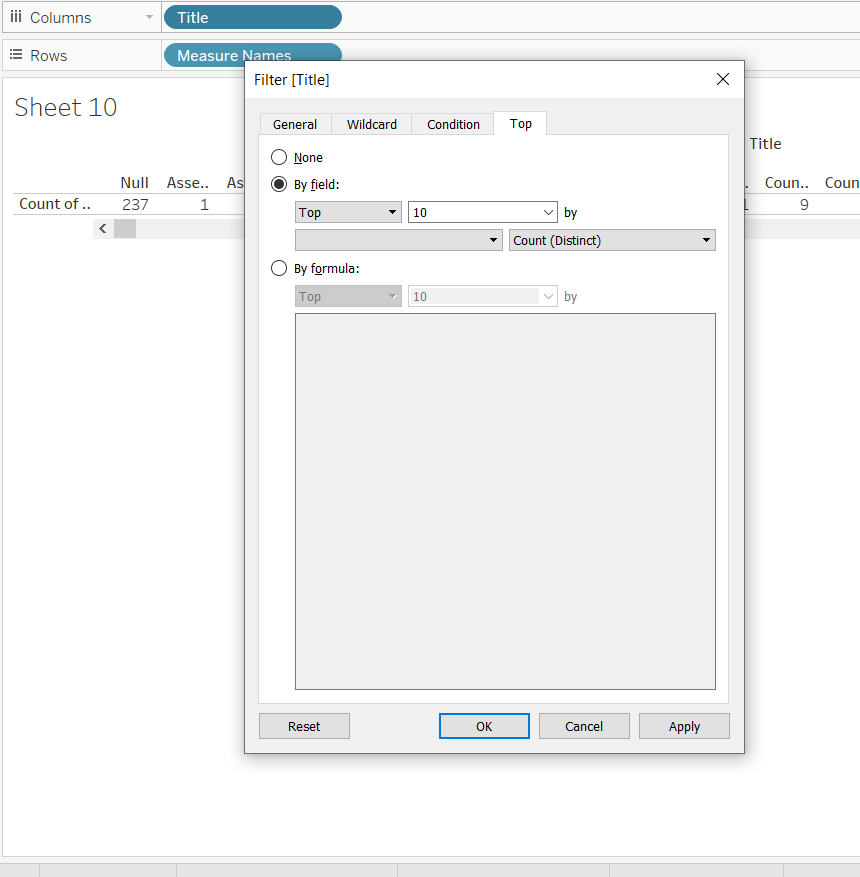
The purpose for the fifth and sixth visual is to compare the changes between the most frequent appeared titles. Now, we make another visualization in which we will make a bar chart using Titles column and the number of records.

1. First, we double tap on the Title and Measure values in order to add them to the columns and rows. Instead of dragging them and choosing as to what in add in column and what to add in Rows, we can always double click and on the fields and let tableau decide.
2. Then, we need to add filters in the Titles section.

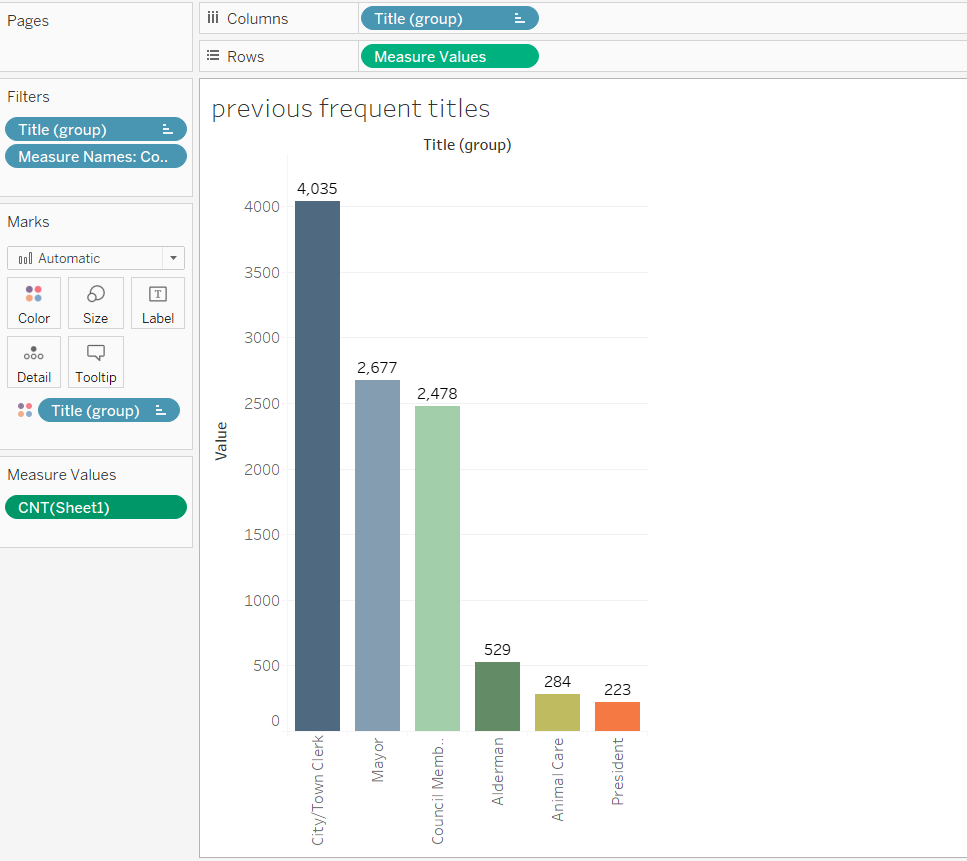




1. Then we need to de-select the Null values, it can make the visualization look bad, if we include all those null values as they would not signify anything.
2. Next step would be going to the Top Bar option and choosing By Field and putting in value. For example, if we want top 6 records by count, we can add values 6. It will automatically filter it out.



1. Then, we can add Title to the Color bar and the visualization is ready. This was done for the previous frequency titles and same can be done for Current frequency titles just by adding the current data and following the same steps.



Step 4: Create Dashboard

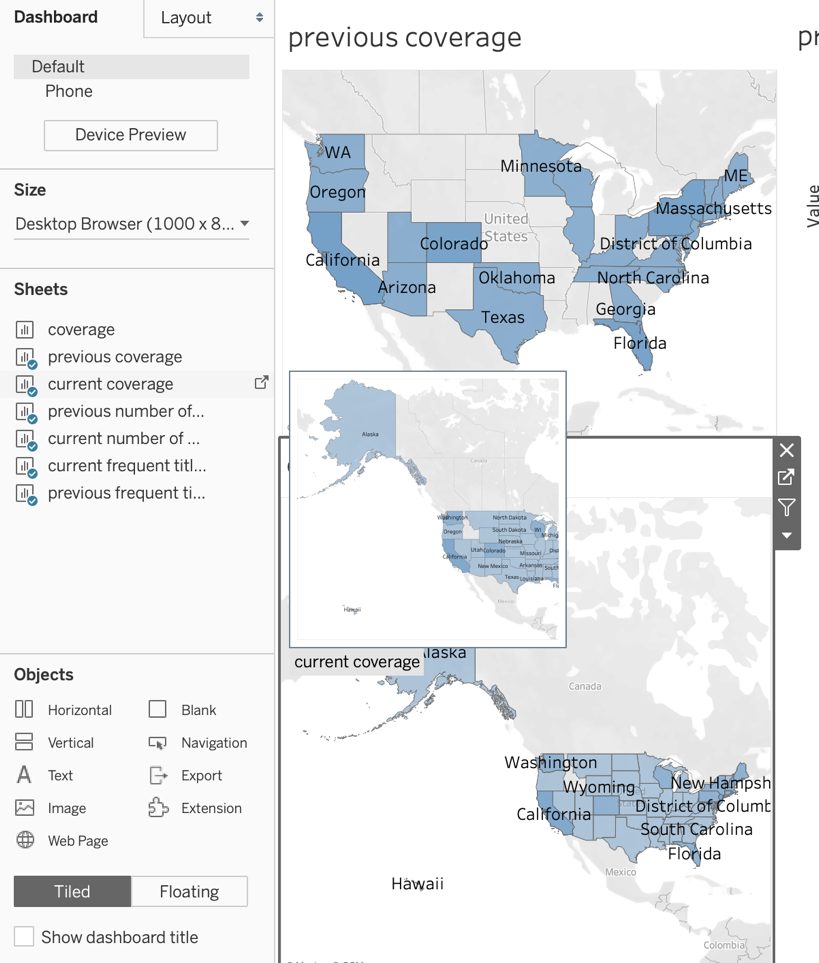
Once you’ve created all visuals on different sheets in Tableau, you can pull them into a dashboard.

1. At the bottom of the workbook, click the New Dashboard icon:

Qr code

Description automatically generated

1. From the Sheets list at left, add views to your dashboard by dragging sheets onto the dashboard at right.



1. For each view added to the dashboard, click on “floating layout” to buy space for multiple visuals.

Graphical user interface, map

Description automatically generated

1. To add interactivity to the dashboard, we used filter option to add actions. For the worksheet we want actions, go to the target worksheet and select actions under the toolbar.

Chart

Description automatically generated

1. Once the Default dialog box is showing, click Add Actions and then select Filter.

Graphical user interface, application, table

Description automatically generated

1. Since we have two data sources for the dashboard, create separate filters that matches with the data resources and the worksheets added to the dashboard. You can also give a name to the filter to differentiate them, and then click “OK” to save actions.

Graphical user interface, application

Description automatically generated

**Conclusion**

Throughout the project, we updated the existing data and added new data points by using web scraping technique and manual way of collecting data. We provided Next Pet contact information of responsible people regarding pet licensing in different cities almost all over the USA. We also used Tableau essential to provide interactive dashboard visualization of our data and its analysis.

Here are some recommendations based on our analysis:

* Explore with different means of data collection methods and use the one that fits your need to most for future contact information collection.
* Due to competitive environment in the pet licensing as some companies have already expand their area of work, we strongly recommend Next Pet start contracting municipalities as soon as possible.
* To get comparative advantage, we recommend Next Pet to offer some no cost services to municipalities such as cloud-based licensing software, implementation planning, program consultations, and community engagement efforts.
* Due to different laws, requirements, and cost of licensing, we recommend Next Pet to create a Licensing Board, which will consist of departments, which will work on executive, legislative as well as statistical issues.
* Beside the data provided by us, we recommend Next Pet to use official website of US Conference of Mayors, where updated contact information of more than 1400 mayors from different cities, which have more than 30000 population, are indicated.

The link of website: <https://www.usmayors.org/mayors/meet-the-mayors/>