

## Restoring Connectivity: Addressing Internet Outages

# Problem Statement



Consumer complaints provide a rich source of data for businesses.



This data can offer valuable insights into customer satisfaction and areas for improvement.



However, the data is often unstructured and messy, making it difficult to analyze.



Our task is to clean, analyze, and visualize this data to uncover actionable insights.

# Business Understanding



Understanding consumer complaints is crucial for businesses.



Analyzing this data can help identify areas that need improvement.



Improvements in these areas can enhance customer satisfaction and loyalty.



Understanding complaints can also lead to product and service innovation.



These actions can increase market share and profitability.

# Data Understanding



The dataset we are working with is sourced from the FCC's Consumer Complaints Data.



It contains valuable information about consumer complaints, including the form (TV, Internet, or Phone), method (Cable, Wireless, Wired, Satellite, or Broadcast), and the specific issue of each complaint.



It provides geographical information about the complainants, including their city, state, and zip code. The dataset spans a significant time period, offering a comprehensive view of consumer complaints trends over time. By analyzing this data, we aim to uncover patterns and trends that can provide actionable insights for businesses.

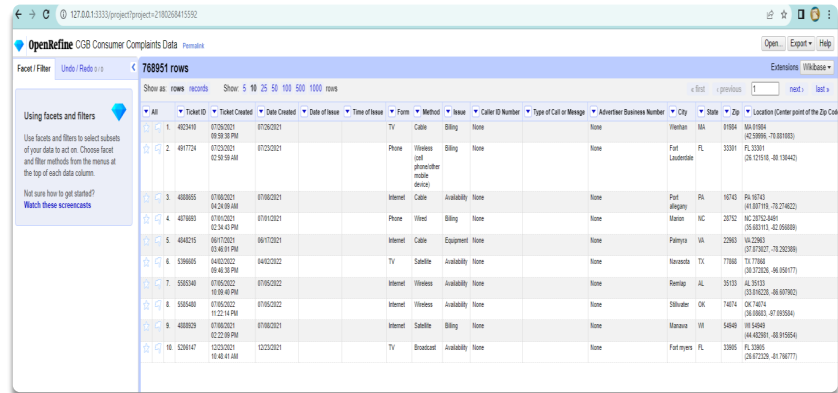
Column Name	Description	Type
Ticket ID	A unique identifier for each ticket or customer service request.	Number
Ticket Created	The date and time when the ticket or customer service request was created.	Date and time (With Timezone)
Date Created	The date when the ticket or customer service request was created (without the timestamp).	Date and time
Form	The method or channel through which the customer submitted the request (e.g., TV, Phone, Internet, Radio, Emergency, Accessibility and Request for Dispute Assistance).	Plain Text
Method	The type of service or technology related to the issue (e.g., Cable, Wireless, Satellite).	Plain Text
Issue	The nature of the problem or reason for the customer service request (e.g., Billing, Availability, Equipment).	Plain Text
City	The city associated with the customer's location or the service issue.	Plain Text
State	The state associated with the customer's location or the service issue (using the two-letter state abbreviation).	Plain Text
Zip	The postal code (ZIP code) associated with the customer's location or service issue.	Plain Text
Location (Center point	The center point coordinates (latitude and longitude) of the ZIP code, representing the approximate geographic location of the	Plain Text

# Data Dictionary

Devendra Bupathi Rapolu

# Data Preprocessing

- ▶ Embarking on our data analysis journey, we first address data cleaning and transformation.
- ▶ We utilize OpenRefine, a powerful tool designed to work with messy data.
- ▶ This step is crucial to ensure the accuracy and reliability of our subsequent analysis.



The screenshot displays the OpenRefine interface for the 'OGB Consumer Complaints Data' project. The main table shows 768,951 rows. The columns are: Ticket ID, Date Created, Date of Issue, Time of Issue, Form, Method, Issue, Caller ID Number, Type of Call or Message, Advertiser Business Number, City, State, and Zip. The table is sorted by Ticket ID in ascending order. The first few rows are visible, showing various complaint details.

	Ticket ID	Date Created	Date of Issue	Time of Issue	Form	Method	Issue	Caller ID Number	Type of Call or Message	Advertiser Business Number	City	State	Zip
1	482349	07/20/2021	07/20/2021	09:39:30 PM	TV	Cable	Billing	None	None	None	Warren	MA	01884
2	4917724	07/20/2021	07/20/2021	02:39:58 AM	Phone	Wireless (cell phone/other mobile device)	Billing	None	None	None	Fort Lauderdale	FL	33301
3	489955	07/18/2021	07/18/2021		Internet	Cable	Availability	None	None	None	Fort Wayne	IN	46740
4	4571693	07/19/2021	07/19/2021		Phone	Wired	Billing	None	None	None	Warren	NC	28782
5	4848215	06/17/2021	06/17/2021		Internet	Cable	Equipment	None	None	None	Palmyra	VA	22963
6	539605	04/02/2022	04/02/2022		TV	Satellite	Availability	None	None	None	Norcross	TX	77066
7	5385540	07/16/2022	07/16/2022		Internet	Wireless	Availability	None	None	None	Remlap	AL	36133
8	5355469	07/16/2022	07/16/2022		Internet	Wireless	Availability	None	None	None	Stillwater	OK	74714
9	4899329	07/18/2021	07/18/2021		Internet	Satellite	Billing	None	None	None	Warren	MI	48090
10	5286447	12/23/2021	12/23/2021		TV	Broadcast	Availability	None	None	None	Fort Myers	FL	33905

Devendra Bupathi Rapolu

# Trimming Whitespace

- ▶ Our initial step in the cleaning process is to trim leading and trailing spaces from the data.
- ▶ This is an essential step as extra spaces can lead to issues during analysis and can result in inaccurate results.
- ▶ By ensuring our data is free from unnecessary spaces, we lay a clean foundation for further steps.

OpenRefine OGB Consumer Complaints Data

Facet/Filter: Undo/Redo

768951 rows

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

	Ticket ID	Date Created	Date of Issue	Time of Issue	Form	Method	Issue	Caller ID Number	Type of Call or Message	Advertiser Business Number	City	State	Zip	Location (Center point of the Zip Code)
1	482349	07/20/2021	07/20/2021	09:39:30 PM	TV	Cable	Billing	None	None	None	Warren	MA	01884	MA 01884 (42.5889, -70.881803)
2	491724	07/20/2021	07/20/2021	02:39:58 AM	Phone	Wireless (cell phone/other mobile device)	Billing	None	None	None	Fort Lauderdale	FL	33301	FL 33301 (26.12191, -80.138442)
3	489955	07/18/2021	07/18/2021	04:29:05 AM	Internet	Cable	Availability	None	None	None	Fort Wayne	IN	46740	IN 46740 (41.037119, -72.274622)
4	457689	07/19/2021	07/19/2021	02:34:42 PM	Phone	Wired	Billing	None	None	None	Warren	NC	28752	NC 28752-4491 (35.887113, -82.958889)
5	4848215	06/17/2021	06/17/2021	03:48:31 PM	Internet	Cable	Equipment	None	None	None	Palmyra	VA	22963	VA 22963 (37.873827, -78.202388)
6	539605	04/02/2022	04/02/2022	09:48:38 PM	TV	Satellite	Availability	None	None	None	Norman	TX	77060	TX 77060 (29.72525, -96.959177)
7	538540	07/16/2022	07/16/2022	10:08:48 PM	Internet	Wireless	Availability	None	None	None	Remlap	AL	36133	AL 36133 (33.245228, -86.957962)
8	535490	07/16/2022	07/16/2022	11:22:14 PM	Internet	Wireless	Availability	None	None	None	Stillwater	OK	74714	OK 74714 (36.9863, -97.69204)
9	489929	07/18/2021	07/18/2021	12:25:05 PM	Internet	Satellite	Billing	None	None	None	Warren	MI	48040	MI 48040 (44.42381, -80.919554)
10	520647	12/23/2021	12/23/2021	10:42:41 AM	TV	Broadcast	Availability	None	None	None	Fort Myers	FL	33905	FL 33905 (26.872229, -81.766777)

Devendra Bupathi Rapolu

# Transforming Data Types

- ▶ Next, we focus on the 'Zip' column, which was initially in text format.
- ▶ We transform this column into a numeric format, a crucial step for any subsequent numerical operations or analyses that we might need to perform.
- ▶ This transformation enhances the versatility of our dataset, allowing for more complex analyses.

OpenRefine CGB Consumer Complaints Data [Permalink](#)

Facet / Filter Undo / Redo 0 / 0 768951 rows

Refresh Reset all Remove all Show as: rows records Show: 5 10

State change 56 choices Sort by: name count Cluster

AK 1475  
AL 10509  
AR 6498  
AS 82  
AZ 17047  
CA 90696  
CO 14118  
CT 8872  
DC 3172  
DE 3312  
FL 72024  
GA 30731

Zip change reset No numeric value present.

	All	Ticket ID	Ticket Created
1.	4923410	07/26/2021 09:59:38 PM	07/26/2021 09:59:38 PM
2.	4917724	07/23/2021 02:50:59 AM	07/23/2021 02:50:59 AM
3.	4888655	07/08/2021 04:24:59 AM	07/08/2021 04:24:59 AM
4.	4876693	07/01/2021 02:34:43 PM	07/01/2021 02:34:43 PM
5.	4848215	06/17/2021 03:46:01 PM	06/17/2021 03:46:01 PM
6.	5396605	04/02/2022 09:46:38 PM	04/02/2022 09:46:38 PM
7.	5585340	07/05/2022 10:09:40 PM	07/05/2022 10:09:40 PM
8.	5585480	07/05/2022 11:22:14 PM	07/05/2022 11:22:14 PM
9.	4888929	07/08/2021 02:22:09 PM	07/08/2021 02:22:09 PM

OpenRefine CGB Consumer Complaints Data [Permalink](#)

Facet / Filter Undo / Redo 1 / 1 768951 rows

Refresh Reset all Remove all Show as: rows records Show: 5 10 25 50 100 500

State change 56 choices Sort by: name count Cluster

AK 1475  
AL 10509  
AR 6498  
AS 82  
AZ 17047  
CA 90696  
CO 14118  
CT 8872  
DC 3172  
DE 3312  
FL 72024  
GA 30731

Zip change reset

0 — 100,000

☒ Numeric ☒ Non-numeric ☒ Blank ☐ Error

	All	Ticket ID	Ticket Created	Date Created
1.	4923410	07/26/2021 09:59:38 PM	07/26/2021 09:59:38 PM	07/26/2021
2.	4917724	07/23/2021 02:50:59 AM	07/23/2021 02:50:59 AM	07/23/2021
3.	4888655	07/08/2021 04:24:59 AM	07/08/2021 04:24:59 AM	07/08/2021
4.	4876693	07/01/2021 02:34:43 PM	07/01/2021 02:34:43 PM	07/01/2021
5.	4848215	06/17/2021 03:46:01 PM	06/17/2021 03:46:01 PM	06/17/2021
6.	5396605	04/02/2022 09:46:38 PM	04/02/2022 09:46:38 PM	04/02/2022
7.	5585340	07/05/2022 10:09:40 PM	07/05/2022 10:09:40 PM	07/05/2022
8.	5585480	07/05/2022 11:22:14 PM	07/05/2022 11:22:14 PM	07/05/2022
9.	4888929	07/08/2021 02:22:09 PM	07/08/2021 02:22:09 PM	07/08/2021
10.	5206147	12/23/2021 10:48:41 AM	12/23/2021 10:48:41 AM	12/23/2021



# Removing Redundancies

- ▶ We then proceed to remove columns that are mostly null or not necessary for our intended analysis.
- ▶ These include 'Date of Issue', 'Time of Issue', 'Caller ID Number', 'Type of Call or Message', 'Advertiser Business Number', and 'Type of Property Goods or Services'.
- ▶ This step helps us to focus on the relevant data and reduces the dimensionality of the dataset, making subsequent analysis more manageable and efficient.

Complaints Data [Permalink](#) Remove column Type of Property Goods or Services Undo

768951 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

	Ticket ID	Ticket Created	Date Created	Form	Method	Issue	City	State	Zip	Location (Center point of the Zip Code)
1.	4023410	07/26/2021 09:59:38 PM	07/26/2021	TV	Cable	Billing	Wenham	MA	1984	MA 01984 (42.59996, -70.881083)
2.	4917724	07/23/2021 02:59:59 AM	07/23/2021	Phone	Wireless (cell phone/other mobile device)	Billing	Fort Lauderdale	FL	33301	FL 33301 (26.121516, -80.130442)
3.	4088655	07/08/2021 04:24:09 AM	07/08/2021	Internet	Cable	Availability	Port Allegany	PA	16743	PA 16743 (41.807119, -78.274622)
4.	4076993	07/01/2021 02:34:43 PM	07/01/2021	Phone	Wireless	Billing	Marion	NC	28752	NC 28752-6491 (35.683113, -82.056689)
5.	4840215	06/17/2021 03:46:01 PM	06/17/2021	Internet	Cable	Equipment	Palmyra	VA	22963	VA 22963 (37.873027, -78.292389)
6.	5396605	04/02/2022 09:46:38 PM	04/02/2022	TV	Satellite	Availability	Navasota	TX	77868	TX 77868 (30.372626, -96.050177)
7.	5585340	07/05/2022 10:09:40 PM	07/05/2022	Internet	Wireless	Availability	Remlap	AL	35133	AL 35133 (33.816228, -86.607902)
8.	5585480	07/05/2022 11:22:14 PM	07/05/2022	Internet	Wireless	Availability	Stillwater	OK	74074	OK 74074 (36.06663, -97.093584)
9.	4888929	07/08/2021 02:22:09 PM	07/08/2021	Internet	Satellite	Billing	Manauva	WI	54949	WI 54949 (44.482501, -88.915654)
10.	5206147	12/23/2021 10:40:41 AM	12/23/2021	TV	Broadcast	Availability	Fort Myers	FL	33905	FL 33905 (26.672329, -81.766777)

# Handling Missing Values

- ▶ Our next step is to handle missing data, specifically focusing on the 'Zip' column.
- ▶ Any rows with blank values in this column are removed from the dataset.
- ▶ Handling missing data is important as they can skew the results and often lead to inaccurate analysis.

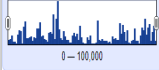
OpenRefine CGB Consumer Complaints Data [Permalink](#)

Facet / Filter Undo / Redo 0/0 766416 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Refresh Reset all Remove all

Zip change reset



0 — 100,000

	All	Ticket ID	Ticket Created	Date Created	Form	Method	Issue	City	State	Zip	Location (Center point of the Zip Code)
1.	4823410		07/26/2021 09:59:38 PM	07/26/2021	TV	Cable	Billing	Wentham	MA	1984	MA 01984 (42.59996, -70.881083)
2.	4917724		07/23/2021 02:59:59 AM	07/23/2021	Phone	Wireless (cell phone/other mobile device)	Billing	Fort Lauderdale	FL	33301	FL 33301 (26.121518, -80.130442)
3.	4888665		07/08/2021 04:24:09 AM	07/08/2021	Internet	Cable	Availability	Port Allegany	PA	16743	PA 16743 (41.807119, -78.274822)
4.	4079893		07/01/2021 02:34:43 PM	07/01/2021	Phone	Wired	Billing	Marion	NC	28752	NC 28752-4491 (35.883113, -82.856889)
5.	4848215		06/17/2021 03:48:01 PM	06/17/2021	Internet	Cable	Equipment	Palmira	VA	22963	VA 22963 (37.873027, -78.282389)
6.	5396895		04/02/2022 09:48:38 PM	04/02/2022	TV	Satellite	Availability	Naruto	TX	77868	TX 77868 (30.372826, -96.056177)
7.	5585340		07/05/2022 10:09:40 PM	07/05/2022	Internet	Wireless	Availability	Remlap	AL	35133	AL 35133 (33.916228, -86.607982)
8.	5305480		07/05/2022 11:22:14 PM	07/05/2022	Internet	Wireless	Availability	Silvinder	OK	74074	OK 74074 (36.89692, -97.893584)
9.	4808829		07/08/2021 02:22:08 PM	07/08/2021	Internet	Satellite	Billing	Manawa	WI	54949	WI 54949 (44.482981, -88.915654)
10.	5206147		12/23/2021 10:48:41 AM	12/23/2021	TV	Broadcast	Availability	Fort Myers	FL	33905	FL 33905 (26.672329, -81.766777)

## Data Extraction for Latitude and Longitude

- ▶ We then turn our attention to the 'Location (Center point of the Zip Code)' column, which contains both latitude and longitude information.
- ▶ We extract and split this data into two new columns, 'Latitude' and 'Longitude', for easier analysis.
- ▶ This step simplifies the data structure and allows for more precise geographical analysis, enhancing the depth and accuracy of our insights.

### Custom text transform on column Location (Center point of the Zip Code)

Expression Language General Refine Expression Language (GREL)  
`value.replace(/[()]/, "")` No syntax error.

Preview <span>History</span> <span>Starred</span> <span>Help</span>		
row	value	value.replace(/[()]/, "")
1.	MA 01984 (42.59996, -70.881083)	MA 01984 42.59996, -70.881083
2.	FL 33301 (26.121518, -80.130442)	FL 33301 26.121518, -80.130442
3.	PA 16743 (41.807119, -78.274622)	PA 16743 41.807119, -78.274622
4.	NC 28752-8491 (25.602112, -83.856000)	NC 28752-8491 25.602112, -83.856000

On error ☒ keep original ☐ Re-transform up to  times until no change  
☐ set to blank  
☐ store error

OK

Cancel

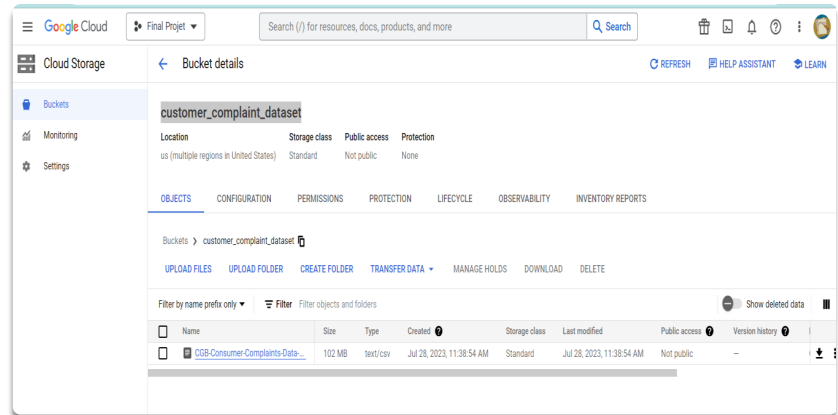
# Exporting Cleaned Data

- ▶ Having completed these preprocessing steps, we are ready to export our cleaned data.
- ▶ The cleaned data is exported into a new CSV file, ready for in-depth analysis and visualization.

A	B	C	D	E	F	G	H	I	J	K	L
Ticket ID	Ticket Created	Date Created	Form	Method	Issue	City	State	Zip	Location (Center point of the Zip Code)	Latitude	Longitude
4923410	7/26/2021 21:59	7/26/2021	TV	Cable	Billing	Wenhan	MA	1984	42.59996, -70.881083	42.59996	-70.881083
4917724	7/23/2021 2:50	7/23/2021	Phone	Wireless (c	Billing	Fort Laude	FL	33301	26.121518, -80.130442	26.121518	-80.130442
4888655	7/8/2021 4:24	7/8/2021	Internet	Cable	Availability	Port allega	PA	16743	41.807119, -78.274622	41.807119	-78.274622
4876693	7/1/2021 14:34	7/1/2021	Phone	Wired	Billing	Marion	NC	28752	35.683113, -82.056889	35.683113	-82.056889
4848215	6/17/2021 15:46	6/17/2021	Internet	Cable	Equipment	Palmyra	VA	22963	37.873027, -78.292389	37.873027	-78.292389
5396605	4/2/2022 21:46	4/2/2022	TV	Satellite	Availability	Navasota	TX	77868	30.372826, -96.050177	30.372826	-96.050177
5585340	7/5/2022 22:09	7/5/2022	Internet	Wireless	Availability	Remlap	AL	35133	33.816228, -86.607902	33.816228	-86.607902
5585480	7/5/2022 23:22	7/5/2022	Internet	Wireless	Availability	Stillwater	OK	74074	36.08683, -97.093584	36.08683	-97.093584
4888929	7/8/2021 14:22	7/8/2021	Internet	Satellite	Billing	Manawa	WI	54949	44.482981, -88.915654	44.482981	-88.915654
5206147	12/23/2021 10:48	12/23/2021	TV	Broadcast	Availability	Fort myers	FL	33905	26.672329, -81.766777	26.672329	-81.766777
4958362	8/12/2021 21:09	8/12/2021	Phone	Wireless (c	Equipment	Evansville	IN	47711	38.011217, -87.5379	38.011217	-87.5379
5206170	12/23/2021 12:50	12/23/2021	Internet	Cable	Availability	District hei	MD	20747	38.85327, -76.887155	38.85327	-76.887155
5396203	4/2/2022 16:44	4/2/2022	Internet	Cable	Availability	Fontana	CA	92335	34.087599, -117.458229	34.087599	-117.458229
4935164	8/1/2021 13:56	8/1/2021	Internet	Cable	Billing	Shelbyville	IN	46176	39.526012, -85.773612	39.526012	-85.773612
4917318	7/22/2021 23:30	7/22/2021	TV	Cable	Billing	Lebanon	PA	17042	40.304366, -76.426501	40.304366	-76.426501
4883372	7/5/2021 23:26	7/5/2021	Internet	Cable	Availability	Pickens	SC	29671	34.92527, -82.714369	34.92527	-82.714369
5583586	7/5/2022 14:09	7/5/2022	Phone	Wireless (c	Billing	Oklahoma	OK	73162	35.581462, -97.641081	35.581462	-97.641081
5583912	7/5/2022 15:56	7/5/2022	TV	Broadcast	Availability	Sherman	TX	75090	33.620647, -96.571184	33.620647	-96.571184
4957792	8/12/2021 18:58	8/12/2021	Emergency	Interference		Algonquin	IL	60102	42.161844, -88.309945	42.161844	-88.309945
4932150	7/30/2021 9:38	7/30/2021	Phone	Wireless (c	Billing	Waterford	PA	16441	41.951558, -79.990381	41.951558	-79.990381
4954346	8/11/2021 12:09	8/11/2021	Phone	Wireless (c	Billing	Oceanside	CA	92054	33.193155, -117.358567	33.193155	-117.358567

# Setting up Google Project and Storage

- ▶ To kickstart our analysis, we create a new Google Cloud project named 'Final\_Project'.
- ▶ We then set up a Google Cloud Storage bucket named 'customer\_complaint\_dataset'.
- ▶ The cleaned dataset is uploaded to this bucket, marking the first step in our cloud-based data analysis journey.



# Creating a Table in BigQuery

- ▶ Next, we move to Google BigQuery, a web service from Google that is used for handling and analyzing big data.
- ▶ We create a table in BigQuery to store our cleaned dataset.
- ▶ With our data now in BigQuery, we are ready to perform fast, SQL-like queries against our multi-terabyte datasets.

The screenshot displays the Google Cloud BigQuery interface. On the left, the 'Analysis' sidebar shows the 'SQL workspace' selected. The 'Explorer' pane in the center shows the project hierarchy, with 'customer\_data' selected under 'customer\_complaint\_dataset'. The main pane on the right shows the 'customer\_data' table schema. The schema table lists the following fields:

Field name	Type	Mode	Key	Collation
<input type="checkbox"/> Ticket_ID	INTEGER	NULLABLE		
<input type="checkbox"/> Ticket_Created	TIMESTAMP	NULLABLE		
<input type="checkbox"/> Date_Created	DATE	NULLABLE		
<input type="checkbox"/> Form	STRING	NULLABLE		
<input type="checkbox"/> Method	STRING	NULLABLE		
<input type="checkbox"/> Issue	STRING	NULLABLE		
<input type="checkbox"/> City	STRING	NULLABLE		
<input type="checkbox"/> State	STRING	NULLABLE		
<input type="checkbox"/> Zip	INTEGER	NULLABLE		
<input type="checkbox"/> Location_Center_point_of_the_Zip_Code	STRING	NULLABLE		
<input type="checkbox"/> Latitude	FLOAT	NULLABLE		
<input type="checkbox"/> Longitude	FLOAT	NULLABLE		

At the bottom of the schema view, there are two buttons: 'EDIT SCHEMA' and 'VIEW ROW ACCESS POLICIES'.

# Distribution of Complaints by Form

- Query: Count of complaints by form
- Description: This pie chart shows the distribution of complaints by form, revealing that TV, Phone and Internet are the most common forms of complaints.

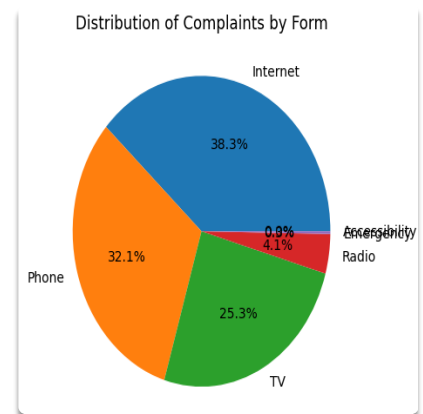
customer\_data x \*Untitled x \*Untitled 2 x

Untitled RUN SAVE SHARE SCHEDULE

```
1 SELECT Form, COUNT(*) as Count
2 FROM `final-projet-394216.customer_complaint_dataset.customer_data`
3 GROUP BY Form
4 ORDER BY Count DESC
```

Query results

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS
Row	Form	Count	
1	Internet	293386	
2	Phone	245658	
3	TV	193722	
4	Radio	31512	
5	Emergency	2137	
6	Accessibility	1	



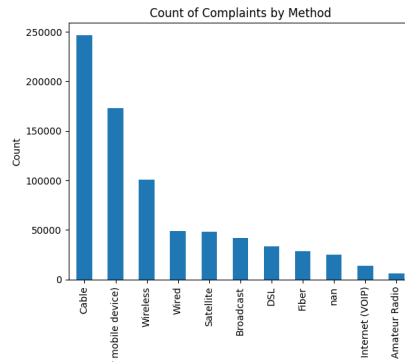
# Distribution of Complaints by Method

Untitled 2

```
1 SELECT Method, COUNT(*) as Count
2 FROM "f5na1-project-394216.customer_complaint_dataset.customer_data"
3 GROUP BY Method
4 ORDER BY Count DESC
5
```

Query results

Row	Method	Count
1	Cable	246760
2	Wireless (cell phone/other mo...	172950
3	Wireless	100795
4	Wired	49048
5	Satellite	48274
6	Broadcast	41901
7	DSL	33337
8	Fiber	28384
9	nan	25071
10	Internet (VOIP)	13863
11	Amateur Radio	5833



- ▶ Query: Count of complaints by method
- ▶ Description: This bar chart displays the distribution of complaints by method, indicating that Cable and Wireless are the most common methods associated with complaints.



## Count of complaints by form and method

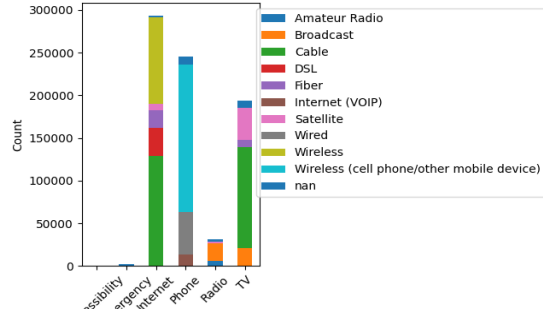
- **Description:** This stacked bar chart shows the distribution of complaints by form and method, revealing which combinations are most common

```
1 SELECT Form, Method, COUNT(*) as Count
2 FROM 'final-projet-394216.customer_complaint_dataset.customer_data'
3 GROUP BY Form, Method
4 ORDER BY Form, Count DESC
```

### Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	Form	Method	Count		
1	Accessibility	nan	1		
2	Emergency	nan	2137		
3	Internet	Cable	128580		
4	Internet	Wireless	100795		
5	Internet	DSL	33337		
6	Internet	Fiber	20432		
7	Internet	Satellite	7949		
8	Internet	nan	2293		
9	Phone	Wireless (cell phone/other mo...	172950		

Count of Complaints by Form and Method



# Top Issues in Complaints

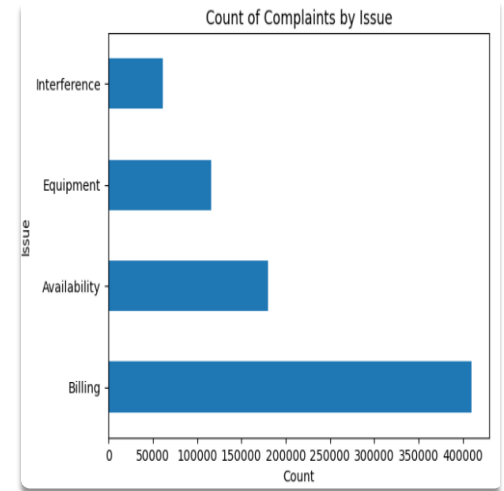
- Query: Top issues in complaints
- Description: This bar chart highlights the top issues in complaints, with Billing and Availability being the most frequent issues.

Untitled 3 [RUN] [SAVE] [SHARE] [SCHEDULE] [MORE]

```
1 SELECT Issue, COUNT(*) as Count
2 FROM 'final-projet-394216.customer_complaint_dataset.customer_data'
3 GROUP BY Issue
4 ORDER BY Count DESC
5
```

Query results

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	EXECUTION G
Row	Issue	Count		
1	Billing	409520		
2	Availability	179854		
3	Equipment	116023		
4	Interference	61019		



# Top 10 Cities with the Highest Number of Complaints

- Query: Top 10 cities with the highest number of complaints
- Description: This bar chart identifies the top 10 cities with the highest number of complaints, providing insights into geographical patterns in the data.

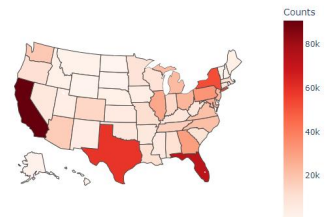
Untitled 4 [RUN](#) [SAVE](#) [SHARE](#)

```
1 SELECT State, COUNT(*) as Count
2 FROM `final-project-394216.customer_complaint_dataset.cust
3 GROUP BY State
4 ORDER BY Count DESC
```

Query results [SAVE RESULT](#)

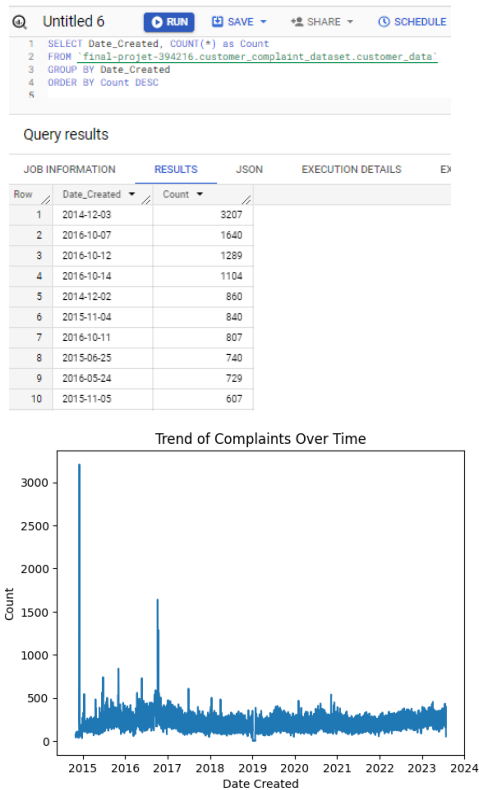
JOB INFORMATION		RESULTS	JSON	EXECUTION DET
Row	State	Count		
1	CA	90671		
2	FL	72004		
3	TX	58598		
4	NY	52163		
5	PA	32972		
6	GA	30708		
7	IL	27821		
8	MD	25912		
9	OH	24180		
10	NJ	23915		

Complaint Count by State



# Monthly Trend of Complaints

- **Query:** Count of complaints by month
- **Description:** This line chart shows the monthly trend of complaints, indicating whether complaints have been increasing or decreasing over time.



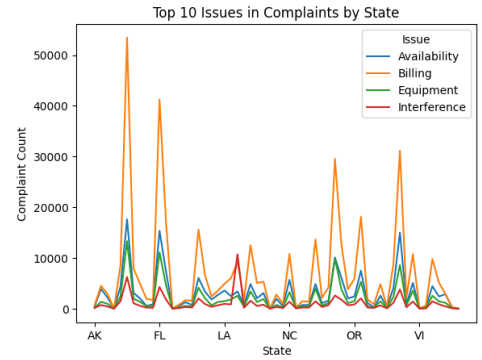
# Top 10 Issues in Complaints by State

- ▶ Query: Top 10 issues in complaints by state
- ▶ Description: This grouped bar chart shows the top 10 issues in complaints by state, providing insights into geographical variations in the issues.

Druvana Sree Sreepada

Query results

Row	State	Issue	Count
1	CA	Billing	53449
2	FL	Billing	41255
3	TX	Billing	31169
4	NY	Billing	29548
5	PA	Billing	18106
6	CA	Availability	17655
7	GA	Billing	16993
8	IL	Billing	15601
9	FL	Availability	15366
10	TX	Availability	15017



# Conclusion



We successfully cleaned and preprocessed a large, messy dataset using OpenRefine, enhancing its usability for analysis.



Our exploration of the dataset using BigQuery and Python revealed interesting insights about consumer complaints, such as the most common issues and their geographical distribution.



The use of various visualization techniques helped us to understand the data better and to communicate our findings effectively.



The project demonstrated the power of cloud-based tools like Google Cloud Storage, Hadoop, and BigQuery in handling and analyzing large datasets.



The insights gained from this project can help businesses to improve their services, address common issues, and enhance customer satisfaction.

# BONUS - Sentiment Analysis Dashboard

**Description:** The Sentiment Analysis Dashboard provides real-time insights into customer sentiments based on their complaints. The interactive dashboard offers a user-friendly interface, allowing businesses to monitor and respond to customer sentiments promptly.

- ▶ **Real-time Sentiment Monitoring:** Live feed of incoming customer complaints with automatic sentiment analysis (positive, negative, neutral).
- ▶ **Sentiment Trends Over Time:** Visualizations (line charts, area plots) showing how customer sentiments change over time, identifying emerging trends.
- ▶ **Word Cloud of Emotions:** Word cloud displaying frequently used words in positive and negative complaints, capturing prevalent emotions.
- ▶ **Sentiment by Product/Service:** Complaint filtering based on specific products or services to understand impact on customer sentiments.
- ▶ **Responding to Negative Sentiments:** Flagging high-impact negative sentiments for prioritized response and mitigation of potential escalations.

# BONUS - Root Cause Analysis for Major Complaints

**Description:** Root Cause Analysis is to identify underlying reasons behind major complaints and recurring issues, enabling targeted solutions and reducing overall complaint volume.

Key Points:

- **Utilize Data Mining Techniques:** Employ association rule mining or decision trees to uncover significant associations or patterns between complaint attributes and their root causes.
  - **Identify Key Contributors:** Discover primary factors contributing to major complaints, including specific services, geographical locations, or demographic segments.
  - **Implement Actionable Recommendations:** Make data-driven decisions to improve processes, enhance customer support, or refine product offerings to mitigate future complaints.
  - **Foster Continuous Improvement:** Regularly perform root cause analysis to instill a culture of continuous improvement, keeping customer satisfaction a top priority.
  - **Close the Feedback Loop:** Integrate insights from root cause analysis into the complaint resolution process, ensuring effective changes are implemented.
- These data-driven techniques enhance the value of the "Analyzing Consumer Complaints: A Data-Driven Approach" project. By leveraging advanced approaches, businesses can gain deeper insights, optimize customer experiences, and drive continuous improvement across their operations.



# References:



<https://opendata.fcc.gov/Consumer/CGB-Consumer-Complaints-Data/3xyp-aqkj/data>



Classroom material / Examples.



Thank you