

FINAL PROJECT REPORT



FOOD WASTE IN DINING SERVICES

ADTA 5240- Harvesting, Storing, and Retrieving Data

Dr. Schenita Floyd

December 6, 2023

Group- 15

Sharath Chandra Reddy Yedavalli

Sushanth Reddy Marreddy

Prem Chand Kothapalli

Meghana Desetty

Hemavalli Gollapudi

Prakash Chiluvuru

Executive Summary

Introduction:

Food waste is a global issue where dining services like university dining halls, hospitals, markets, and restaurants are significant contributors and have a huge amount of food waste generated. This uneaten food generation has not only an impact on the environment but also has financial loss with huge amounts of investments in it (Leal Filho et al., 2023). However, this project aims to study the issue and implement necessary steps to prevent food wastage further in the dining services revolves around handling the eight stages of the data life cycle: Generation, Collection, Processing, Storage, Management, Analysis, Visualization, and Interpretation (HBS Online, 2021). Each stage plays a crucial role in extracting meaningful insights from the raw data collected from the publicly available sites and contributing to informed decision-making processes.

Body (Structure):

Static and Streaming datasets are generated from various sources in the U.S., and they are readily available through websites or links such as data.gov and insights engine ReFED, ensuring the authenticity and reliability of the information. These datasets are collected from the website as a csv file and used for further analysis of food wastage.

The data collected is processed using the Open Refine tool, which offers a versatile set of tools where the data can be cleaned for null values, data inconsistency, changing formats, remove unwanted columns etc. (Ramírez-Gallego et al., 2017). This phase ensures that the data is in structured and usable format. This processed and cleaned data can be used further by the next phases of data lifecycle.

In the storage phase, a project is created by the name “ADTA5240F23G15”. After the creation of the project, the bucket is created along with the clusters and the dataset files are stored in the bucket. The clusters can be managed with the help of manager and worker nodes. The manager node can be used to run the Hadoop system for Hive and Spark analysis for the dataset files stored in the bucket. BigQuery can also be used to manage and analyze the data with its inbuilt capabilities such as ML, and business intelligence (Google Cloud, n.d.). The management stage also involves version control, ensuring governance, orchestrating the data flow and seamless accessibility for downstream process.

With the help of Hadoop commands, Hive and Spark tools enable the execution of the complex queries, with a focus on runtime efficiency. By creating tables in Hive and Spark This helps to extract appropriate results by running SQL queries easily and faster with it on both platforms, required data can be taken using SQL queries.

Conclusion:

Based on the above generated results from Hive and Spark analysis, it can be known that Hive took more time than Spark. However, Hive can be used for performing data analytics on large sets of data, whereas Spark can be used for running big data analytics with a faster approach and time (Berman, n.d.). Furthermore, this analysis can be used by data scientists and data analytics for visualization and interpreting results. Based on these, data scientists and analysts can enhance the future impact on food waste and analyze more efficient solutions to it.

Data Generation and Collection:

Static Dataset Reference and link: Vermont Center for Geographic Information.

(2023, September 15). Food Scrap Generators. *DEC, Waste Management and Prevention Division*. <https://catalog.data.gov/dataset/food-scrap-generators-caa50>

Website link: Data.gov - <https://data.gov/>

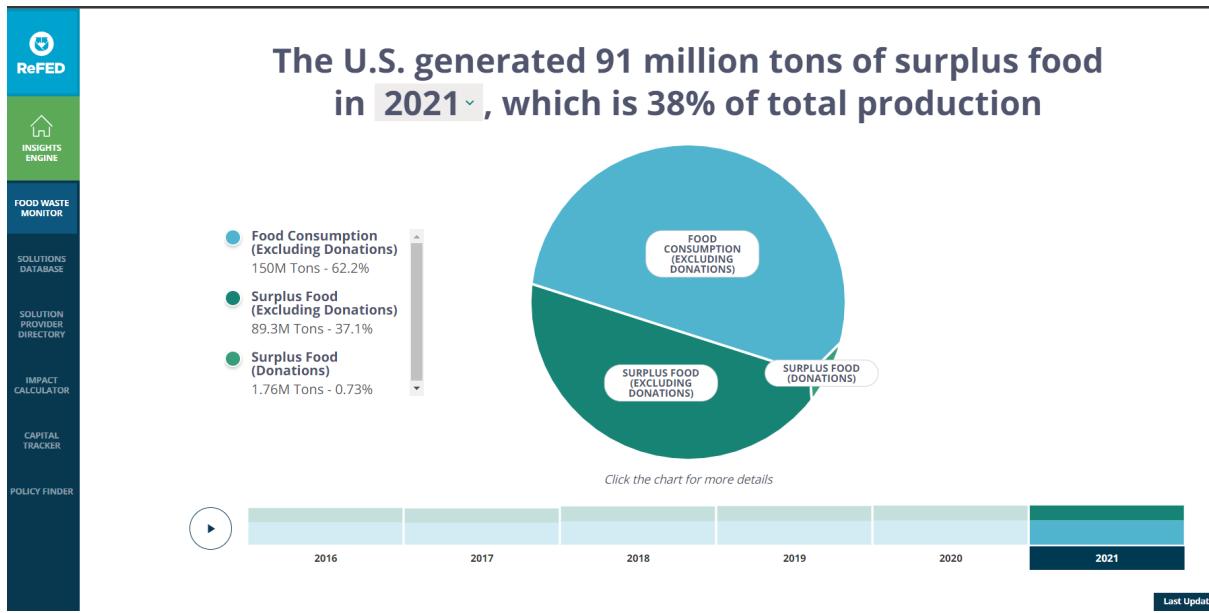
This dataset is collected from data.gov. This dataset has food waste data collected from various sources such as hospitals, restaurants, schools, colleges, etc. in the state of Vermont.

The screenshot shows the Data.gov homepage with a navigation bar at the top featuring the Data.gov logo, DATA, REPORTS, OPEN GOVERNMENT, and CONTACT. Below the header is a blue banner with the text "DATA CATALOG". The main content area displays a dataset titled "Food Scrap Generators" from the "Vermont Center for Geographic Information". On the left, there's a sidebar with links to "read more", "Topics", "Local Government", "Publisher", and "DEC, Waste Management and Prevention Division". The main content area includes a note about non-federal terms of use, a brief description of the dataset, and an "Access & Use Information" section.

Streaming Dataset 1 Reference and link: ReFED Insights Engine. (n.d.). *Food waste monitor*. <https://insights-engine.refed.org/food-waste-monitor?view=overview&year=2021>

Website link: insights-engine.refed.org - https://insights.refed.org/?_ga=2.91862170.1051283044.1701466352-538034373.1701466352

This dataset is taken from the ReFED insights engine which monitors the food waste generated and the cause of wastage in the U.S.



Streaming Dataset 2 Reference and link: Town of Cary, North Carolina. (2023, November 24). Food Waste Recycling Impact. *Cary*.
<https://catalog.data.gov/dataset/food-waste-recycling-impact>

Website link: Data.gov - <https://data.gov/>

This dataset is extracted from the data.gov website where it collects real-time data of food waste from all the sources in the Town of Cary, North Carolina state. This dataset is updated every month for food wastage collection.

The screenshot shows the Data.gov interface. At the top, there's a navigation bar with the Data.gov logo, followed by links for DATA, REPORTS, OPEN GOVERNMENT, and CONTACT. Below this is a blue header bar with the text "DATA CATALOG" and navigation links for HOME / Datasets / Organizations. The main content area displays a dataset for "Town of Cary, North Carolina / Cary". The dataset title is "Food Waste Recycling Impact". It includes a note: "This is a Non-Federal dataset covered by different Terms of Use than Data.gov. See Terms". The dataset was last updated on November 24, 2023. A detailed description follows: "This dataset contains impact report metrics from the Town of Cary's Food Waste Recycling Drop-Off program which began on February 7, 2022. Cary's pilot food waste recycling drop-off is a local option to turn food waste into compost. The site is located at the Citizen's Convenience Center, 313 N. Dixon Ave., and accepts a variety of materials; open to residents, no commercial interests. All material is processed by a local facility and some of that finished compost will be returned to the earth at Good Hope Farm. The dataset is updated monthly." On the left side, there's a sidebar with sections for Topics (Local Government), Publisher (Town of Cary, North Carolina), and Contact (<Nobody>).

Data Processing:

Performed data processing like cleaning the dataset and removing unwanted rows and columns using OpenRefine for both static and streaming datasets.

Static Dataset: Arranging all the columns with ‘ID’ as the starting column. Changing the ‘ID’, Zip, TonsPerWeek, X, Y, Active’ columns are transformed from text to number.

ID	X	Y	FSGName	TonsPerWeek	Type1	Type2	Type3	Address	Town	County
1.	507300.303655109	106026.651106194	BROOKHAVEN CATERER	0.05	Food Establishment	Caterer		1470 ROUTE 44	WINDSOR	WINDSOR
2.	508517.092486237	110238.699568204	VERMONT WHOLE FOODS CATERING	0.05	Food Establishment	Caterer		153 PEMBROKE RD	WINDSOR	WINDSOR
3.	508517.092486237	110238.699568204	JUNIPER HILL INN	0.735	Food Establishment	Restaurant		153 PEMBROKE ROAD	WINDSOR	WINDSOR
4.	509053.226834364	108528.483279731	WINDSOR ELKS BPO LODGE 1665	1.68	Food Establishment	Restaurant		156 MAIN ST	WINDSOR	WINDSOR
5.	508685.55860496	111516.571731314	JOHN P LARKIN COUNTRY CLUB	0.945	Food Establishment	Restaurant		1571 ROUTE 5 NORTH	WINDSOR	WINDSOR
6.	506317.420927535	109836.322131843	GREEN MOUNTAIN FLOUR	0.05	Food Establishment	Caterer		17 MOONRISE HILL	WINDSOR	WINDSOR
7.	509061.496607626	108395.17647465	Cumberland Farms	0.18	Food Establishment	Caterer		183 MAIN ST	WINDSOR	WINDSOR
8.	508818.676136384	108472.62304178	THIS IS THAT TRADING POST TRADING AS TLC	0.05	Food Establishment	Caterer		19 1/2 UNION ST	WINDSOR	WINDSOR
9.	508818.676136384	108472.62304178	VITTE STOP BBQ	0.05	Food Establishment	Caterer		19 1/2 UNION STREET	WINDSOR	WINDSOR
10.	508526.924901937	108861.091440225	WINDSOR HIGH SCHOOL	0.04795	Food Establishment	School		19 ASCUTNEY ST	WINDSOR	WINDSOR
11.	508767.124222613	110783.40367861	GRANNY O COOKIES	0.13	Food Establishment	Bakery		19 HUNTINGTON DRIVE	WINDSOR	WINDSOR
12.	508470.126830772	108994.337789621	GREEN MOUNTAIN ASIAN MARKET	0.18	Food Establishment	Grocery		2 ASCUTNEY ST	WINDSOR	WINDSOR
13.	509050.434552743	110606.020730204	STUBBS & LAURAS	0.26	Food Establishment	Restaurant		2066 ROUTE 5 N	WINDSOR	WINDSOR
14.	509050.434552743	110606.020730204	FRAZERS PLACE	0.42	Food Establishment	Restaurant		2066 VT RT 5 N	WINDSOR	WINDSOR
15.	509498.076964298	108617.97440312	OUR NEIGHBORHOOD KITCHEN	0.05	Food Establishment	Caterer		27 JARVIS ST	WINDSOR	WINDSOR
16.	509343.034845065	109573.203760249	PRICE CHOPPER STORE #177	0.72	Food Establishment	Grocery		2725 US ROUTE 5 N	WINDSOR	WINDSOR
17.	507854.478021386	109693.511487252	MOUNT ASCUTNEY HOSPITAL & HEALTH CTR	0.3465	Food Establishment	Hospital		289 COUNTY RD	WINDSOR	WINDSOR
18.	507854.478021386	109693.511487252	MT ASCUTNEY HOSPITAL	0.18375	Hospital			289 COUNTY ROAD	WINDSOR	WINDSOR
19.	507891.405948811	112693.219421738	AMERICAN CRAFTED SPIRITS INCORPORATED	0.2625	Food Establishment	Restaurant		3 ARTISANS WAY	WINDSOR	WINDSOR
20.	508394.61651464	111160.664385897	BOSTON DREAMS CATERING LLC	0.05	Food Establishment	Home-based Business		321 TAYLOR DR	WINDSOR	WINDSOR

Creating clusters for the columns ‘FSGName’ and ‘County’, and all clusters with similar names are transformed into one and filling down the values to store in blank cells.

County	Zip	Phone	Active
WINDSOR	5089	(802) 674-5273	1
WINDSOR	5089	8026749966	1
WINDSOR	5089	8026746491	1
WINDSOR	5089	null	1
WINDSOR	5089	5082704407	1
WINDSOR	5089	(802) 674-6776	1
WINDSOR	5089	null	1
WINDSOR	5089	8026746344	1
WINDSOR	5053	null	1
WINDSOR	5089	null	1
WINDSOR	5089	(802) 674-5715	1
WINDSOR	5030	8026749859	1
WINDSOR	5089	(802) 345-8360	1
WINDSOR	5089	8026743470	1
WINDSOR	5089	8026746711	1
WINDSOR	5089	(802) 674-6711	1
WINDSOR	5031	3392220987	1
WINDSOR	5089	0823562819	1
WINDSOR	5089	8026745491	1
WINDSOR	5089	8026742301	1
WINDSOR	5089	(802) 674-6733	1

Facet / Filter Undo / Redo 248 / 249

Refresh Reset all Remove all

FSName change

5503 choices Sort by: name count Cluster

CRISANVER HOUSE 1
CRISIS STABILIZATION & INPATIENT DIVERSION PROGRAM (CSID) 1
CROSSETT BROOK MIDDLE SCHOOL 1
CROSSROADS DELI & BAKERY 1
CROSSROADS B & B 1
CROSSROADS BAR AND GRILL 1
CROSSROADS MOBILE 1
CROSSROADS TRADING POST 1
CROSSROADS TRAVEL CENTER 1
CROWN BAG & GRILL 1

6226 rows

Cluster and edit column "County"

Find groups of different cell values that might be other representations of the same thing. For example, "New York" and "new york" likely refer to the same concept and just differ by capitalization, and "Gödel" and "Gode!" probably refer to the same person. [Find out more...](#)

Method Key collision Keying function Daitch-Mokotoff

12 clusters found

Cluster size	Row count	Values in cluster	Merge?	New cell value	# Choices in cluster
3	46	POULTNEY (33 rows) BOLTON (10 rows) WALDEN (3 rows)	<input type="checkbox"/>	POULTNEY	2 — 3
3	23	CRAFTSBURY (14 rows) SHREWSBURY (7 rows) CRAFTSBURY COMMON (2 rows)	<input type="checkbox"/>	CRAFTSBURY	# Rows in cluster
2	4	WEST BURKE (3 rows) WEST BERKSHIRE	<input type="checkbox"/>	WEST BURKE	Average length of choices
2	9	WEST WINDSOR (8 rows) WEST TOWNSHEND	<input type="checkbox"/>	WEST WINDSOR	Length variance of choices
2	4	SOUTH BARRE (3 rows) SUDSBURY	<input type="checkbox"/>	SOUTH BARRE	
2	25	FAYSTON (15 rows) WESTON (10 rows)	<input type="checkbox"/>	FAYSTON	
2	11	PLAINFIELD (10 rows)	<input type="checkbox"/>	PLAINFIELD	

Select all Deselect all Export clusters Merge selected & re-cluster Merge selected & Close Close

20. 508394.61651464 111160.664385897 21 BOSTON DREAMS CATERING LLC 0.05 Food Establishment Home-based Business 321 TAYLOR DR WINDSOR 5089 1

21. 507956.286665811 112526.647412444 22 HARPOON BREWERY 1.8375 Food Establishment Restaurant 338 RUTH CARNEY DR WINDSOR 5089 1

22. 509295.005548735 109206.512448289 23 NIDAS 0.05 Food Establishment Home-based Business 35 N MAIN ST WINDSOR 5089 1

23. 509181.901911009 109095.258923538 24 HISTORIC HOMES OF RUNNEMEDE 0.0735 Catering/Assisted Living Residential Facility 40 MAXWELL PERKINS LANE WINDSOR 5089 1

Removed unwanted columns like Type3, Phone, Website, and Facebook from the dataset.

Foodwaste_static - OpenRefill Foodwaste_static - OpenRefill +

127.0.0.1:3333/project?project=1798508576427

Open... Export Help

OpenRefill Foodwaste_static Permalink

Facet / Filter Undo / Redo 12 / 12

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 screenshots](#)

6222 records

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

Extensions Wikibase

ID	X	Y	FSName	TonsPerWeek	TYPE1	TYPE2	Address	Town	County	Zip	Active
1.	1	507300.303655109	106026.651106194	BROOKHAVEN CATERER	0.05	Food Establishment	Caterer	1470 ROUTE 44	WINDSOR	WINDSOR	5089 1
2.	2	508517.092486237	110238.699568204	VERMONT WHOLE FOODS CATERING	0.05	Food Establishment	Caterer	153 PEMBROKE RD	WINDSOR	WINDSOR	5089 1
3.	3	508517.092486237	110238.699568204	JUNIPER HILL INN	0.735	Food Establishment	Restaurant	153 PEMBROKE ROAD	WINDSOR	WINDSOR	5089 1
4.	4	509053.226834364	108528.483279731	WINDSOR ELKS BPO LODGE 1665	1.68	Food Establishment	Restaurant	156 MAIN ST	WINDSOR	WINDSOR	5089 1
5.	5	508685.55860496	111316.571731314	JOHN P LARKIN COUNTRY CLUB	0.945	Food Establishment	Restaurant	1571 ROUTE 5 NORTH	WINDSOR	WINDSOR	5089 1
6.	6	509317.420927535	109836.322131845	GREEN MOUNTAIN FLOUR	0.05	Food Establishment	Caterer	17 MOONRISE HILL	WINDSOR	WINDSOR	5089 1
7.	7	509061.496607626	108395.176478101	Cumberland Farms	0.18	Food Establishment	Caterer	183 MAIN ST	WINDSOR	WINDSOR	5089 1
8.	8	508818.676136384	108472.62304178	THIS N THAT TRADING POST TRADING AS TLC	0.05	Food Establishment	Caterer	19 1/2 UNION ST	WINDSOR	WINDSOR	5089 1
9.	9	508818.676136384	108472.62304178	VITTE STOP BBQ	0.05	Food Establishment	Caterer	19 1/2 UNION STREET	WINDSOR	WINDSOR	5089 1
10.	10	508526.924901937	108861.091440223	WINDSOR HIGH SCHOOL	0.04795	Food Establishment	School	19 ASCUTNEY ST	WINDSOR	WINDSOR	5089 1
11.	11	508767.124222613	110783.403678761	GRANNY O COOKIES	0.13	Food Establishment	Bakery	194 HUNTINGTON DRIVE	WINDSOR	WINDSOR	5053 1
12.	12	508470.126830772	108994.337769821	GREEN MOUNTAIN ASIAN MARKET	0.18	Food Establishment	Grocery	2 ASCUTNEY ST	WINDSOR	WINDSOR	5089 1
13.	13	509050.434552743	110606.020730204	STUBS & LAURAS	0.26	Food Establishment	Restaurant	2066 ROUTE 5 N	WINDSOR	WINDSOR	5089 1
14.	14	509850.434552743	110606.020730204	FRAZERS PLACE	0.42	Food Establishment	Restaurant	2066 VT RT 5 N	WINDSOR	WINDSOR	5030 1
15.	15	509498.076964298	108617.97440312	OUR NEIGHBORHOOD KITCHEN	0.05	Food Establishment	Caterer	27 JARVIS ST	WINDSOR	WINDSOR	5089 1
16.	16	509343.034845065	109573.203760249	PRICE CHOPPER STORE #177	0.72	Food Establishment	Grocery	2725 US ROUTE 5 N	WINDSOR	WINDSOR	5089 1
17.	18	507854.478021386	109693.511487252	MOUNT ASCUTNEY HOSPITAL & HEALTH CTR	0.3465	Food Establishment	Hospital	289 COUNTY RD	WINDSOR	WINDSOR	5089 1
18.	19	507854.478021386	109693.511487252	MT ASCUTNEY HOSPITAL	0.18375	Hospital		289 COUNTY ROAD	WINDSOR	WINDSOR	5089 1
19.	20	507891.405948811	112693.219421738	AMERICAN CRAFTED SPIRITS INCORPORATED	0.2625	Food Establishment	Restaurant	3 ARTISANS WAY	WINDSOR	WINDSOR	5031 1
20.	21	508394.61651464	111160.664385897	BOSTON DREAMS CATERING LLC	0.05	Food Establishment	Home-based Business	321 TAYLOR DR	WINDSOR	WINDSOR	5089 1

Changed all the null zip codes to specific value based on their county.

Facet / Filter Undo / Redo 62 / 62

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

Extensions Wikibase ▾

Refresh Reset all Remove all

Zip change reset
9 choices Sort by: name count Cluster
5456 1
5472 1
5495 1
5602 1
5702 1
5752 1
5753 119
5754 1
5769 1
Facet by choice counts

Zip change reset
1,100 – 6,000
 Numeric Non-numeric Blank Error

County change invert reset
349 choices Sort by: name count
ENOSBURG 35
Enosburg 3
Essex 6
ESSEX 147
FAIR 29
Fair 1
FAIRFAX 14
FAIRFIELD 10
Fairlee 3
FAIRLEE 28
Falls 3
FALLS 53

127 matching rows (6226 total)

All	ID	FSName	TonsPerWeek	Type1	Type2	Address	County	Zip	Phone	Active
1207.	446096.980930901	170628.256573188	1384	AQUA VITEA	0	Food Manufacturer	MIDDLEBURY	5753	802-453-8590	1
1208.	446387.720403124	168302.228076965	1438	TASTE OF INDIA	0.756	Food Establishment	MIDDLEBURY	5753	8023884856	1
1209.	447273.79844614	167795.081590523	1439	SUGAR MAPLE FARMS	0.05	Food Establishment	Caterer	1 SWANAGE CT	MIDDLEBURY	5753 (802) 388-7745
1210.	446653.858264316	168488.972755163	1440	MIDDLEBURY NATURAL FOODS COOPERATIVE	0.72	Food Establishment	Grocery	9 WASHINGTON ST	MIDDLEBURY	5753 802887276
1211.	446766.383855041	168521.400593959	1442	GREEN PEPPERS RESTAURANT	0.525	Food Establishment	Restaurant	10 WASHINGTON ST	MIDDLEBURY	5753 802-388-3164
1212.	446196.85116021998	166525.922444225	1443	EASTVIEW AT MIDDLEBURY	0.3255	Nursing/Accommodated Living	Residential Facility	100 EASTVIEW TERRACE	MIDDLEBURY	5753 805-989-7500
1213.	446508.628652386	168379.030618539	1444	SABA SABA THAI CUISINE	0.575	Food Establishment	Restaurant	22 MERCHANTS ROW	MIDDLEBURY	5602 8029897376
1214.	446702.237203554	168521.917395556	1445	MIDDLEBURY BAGEL & DELICATESSEN	0.2625	Food Establishment	Restaurant	11 WASHINGTON ST	MIDDLEBURY	5753 8024622230
1215.	446285.64042221	168569.3551805	1446	STONE LEAF TEAHOUSE LLC	0.21	Food Establishment	Restaurant	111 MELLETT ST	MIDDLEBURY	5753 null
1216.	446383.76030063	168624.417707651	1447	PORTER HOSPITAL	0.2825	Food Establishment	Hospital	115 PORTER DR	MIDDLEBURY	5753 8023884775
1217.	448719.398628531	162883.5448587363	1448	HEART TO HEARTH	0.05	Food Establishment	Caterer	1205 THREE MILE BRIDGE ROAD	MIDDLEBURY	5753 null
1218.	450788.281894946	169123.596361847	1449	ALFRESCO TENTS CATERING AND FLOWERS	0.05	Food Establishment	Caterer	121 MUNGER ST	MIDDLEBURY	5753 null
1219.	446519.763479317	170745.717043813	1450	WHIRLIES WORLD	0.05	Food Establishment	Caterer	1232 EXCHANGE STREET	MIDDLEBURY	5752 null
1220.	445948.99358181	167727.987713034	1451	MIDDLEBURY COLLEGES	1.42154	School	College/University	131 SOUTH MAIN STREET	MIDDLEBURY	5753 null
1221.	446293.302548403	168536.3373242933	1452	AMERICAN FLATBREAD	1.0395	Food Establishment	Restaurant	137 MAPLE ST STE 29F	MIDDLEBURY	5753 8023883300
1222.	446293.302548403	168536.3373242933	1453	MOONIES DELI	0.338	Food Establishment	Restaurant	137 MAPLE ST	MIDDLEBURY	5753 8025225375
1223.	446266.538452999	168192.095652124	1454	OTTER CREEK DELI & BAKERY	0.0525	Food Establishment	Restaurant	14 COLLEGE ST	MIDDLEBURY	5753 8023883371
1224.	446597.819405008	168500.536759078	1455	MIDDLEBURY INN	3.19	Food Establishment	Lodging	14 COURT SQ	MIDDLEBURY	5753 802-388-4961
1225.	447887.008941454	165771.072336654	1456	KINGS PLATE	0.26	Food Establishment	Restaurant	1499 ROUTE 7 S	MIDDLEBURY	5753 802-388-2792
1226.	448830.008941454	165771.072336654	1457	MAGIC CHINESE WOK INC	0.525	Food Establishment	Restaurant	1499 US ROUTE 7	MIDDLEBURY	5753 2129331680
1227.	446448.890727616	168923.98277467	1458	ELM STREET GROUP HOME	0.02625	Nursing/Accommodated Living	Residential Facility	15 ELM STREET	MIDDLEBURY	5753 (802) 388-4021
1228.	446014.233682933	170883.158496214	1459	BULMERS AMERICA	0	Food Manufacturer	Alcohol Manufacturer	153 POND LN	MIDDLEBURY	5753 (802) 388-700
1229.	446014.233682933	170883.158496214	1460	GREEN MOUNTAIN BAKED	2.402884615	Food Manufacturer	Food Processor	1321 EXCHANGE ST	MIDDLEBURY	5753 8023880700
1230.	448877.795454888	165728.856992217	1461	A & W	0.26	Food Establishment	Caterer	1557 US ROUTE 7 S	MIDDLEBURY	5753 8023882876
1231.	446659.994307758	168255.795688868	1462	MIDDLEBURY VILLAGE DEPOT	0.126	Food Establishment	Restaurant	16 COURT ST	MIDDLEBURY	5753 8023883000
1232.	445710.988686505	169874.491729114	1463	VS	0.13	Food Establishment	Bakery	161 SEYMOUR STREET EXT	MIDDLEBURY	5753 null
1233.	449020.987701714	165570.19149036	1464	DINKY DONUTS	0.05	Food Establishment	Caterer	1663 ROUTE 7 S	MIDDLEBURY	5753 (802) 839-5099
1234.	451731.010104048	166461.006508047	1465	ELIZABETH'S CREATIVE CREATIONS	0.05	Food Establishment	Caterer	1676 CASE STREET	MIDDLEBURY	5753 null
1235.	452005.89309353	166739.015723258	1466	BABAS CATERING	0.05	Food Establishment	Home-based	1698 CASE ST	MIDDLEBURY	5753 8023730456

Renamed the column names for easy identification.

Foodwaste_static - OpenRefine Foodwaste_static Permalink

Remove column Town Undo

Open... Export Help

Facet / Filter Undo / Redo 17 / 17

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

6222 records

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

Extensions Wikibase ▾

All	ID	Latitude	Longitude	FSName	TonsPerWeek	Type	Sub_Type	Address	County	Zip	Active
1.	1	507300.303655109	10626.651106194	BROOKHAVEN CATERER	0.05	Food Establishment	Caterer	1470 ROUTE 44	WINDSOR	5089	1
2.	2	508517.092486237	110238.69568204	VERMONT WHOLE FOODS CATERING	0.05	Food Establishment	Caterer	153 PEMBROKE RD	WINDSOR	5089	1
3.	3	508517.092486237	110238.69568204	JUNIPER HILL INN	0.735	Food Establishment	Restaurant	153 PEMBROKE ROAD	WINDSOR	5089	1
4.	4	509053.226834364	108528.483279731	WINDSOR ELKS BPO LODGE 1665	1.68	Food Establishment	Restaurant	156 MAIN ST	WINDSOR	5089	1
5.	5	508685.55860496	111316.571731314	JOHN P LARKIN COUNTRY CLUB	0.945	Food Establishment	Restaurant	1571 ROUTE 5 NORTH	WINDSOR	5089	1
6.	6	508317.420927535	109836.322131845	GREEN MOUNTAIN FLOUR	0.05	Food Establishment	Caterer	17 MOONRISE HILL	WINDSOR	5089	1
7.	7	509061.496607626	108395.176478101	Cumberland Farms	0.18	Food Establishment	Caterer	183 MAIN ST	WINDSOR	5089	1
8.	8	508818.676136384	108472.62304178	THIS IS NOT TRADING POST TRADING AS TLC	0.05	Food Establishment	Caterer	19 1/2 UNION ST	WINDSOR	5089	1
9.	9	508818.676136384	108472.62304178	VITTE STOP BBQ	0.05	Food Establishment	Caterer	19 1/2 UNION STREET	WINDSOR	5089	1
10.	10	508526.924901937	108861.091440225	WINDSOR HIGH SCHOOL	0.04795	Food Establishment	School	19 ASCUTNEY ST	WINDSOR	5089	1
11.	11	508767.124222613	110783.403768767	GRANNY O COOKIES	0.13	Food Establishment	Bakery	194 HUNTINGTON DRIVE	WINDSOR	5053	1
12.	12	508470.126830772	108994.33778621	GREEN MOUNTAIN ASIAN MARKET	0.18	Food Establishment	Grocery	2 ASCUTNEY ST	WINDSOR	5089	1
13.	13	509050.943552743	110606.020730204	STUBS & LAURAS	0.26	Food Establishment	Restaurant	2066 ROUTE 5 N	WINDSOR	5089	1
14.	14	509050.943552743	110606.020730204	FRAZERS PLACE	0.42	Food Establishment	Restaurant	2066 VT RT 5 N	WINDSOR	5030	1
15.	15	509498.076964298	108617.97440312	OUR NEIGHBORHOOD KITCHEN	0.05	Food Establishment	Caterer	27 JARVIS ST	WINDSOR	5089	1
16.	16	509343.034845065	109573.203760249	PRICE CHOPPER STORE #177	0.72	Food Establishment	Grocery	2725 US ROUTE 5 N	WINDSOR	5089	1
17.	17	507854.478021386	109693.511487252	OUNT ASCUTNEY HOSPITAL & HEALTH CTR	0.3465	Food Establishment	Hospital	289 COUNTY RD	WINDSOR	5089	1
18.	18	507854.478021386	109693.511487252	MT ASCUTNEY HOSPITAL	0.18375	Hospital	289 COUNTY ROAD	WINDSOR	5089	1	
19.	19	507891.405948811	112693.219427178	AMERICAN CRAFTED SPIRITS INCORPORATED	0.2625	Food Establishment	Restaurant	3 ARTISANS WAY	WINDSOR	5031	1
20.	20	508394.61651464	111160.664385897	BOSTON DREAMS CATERING LLC	0.05	Food Establishment	Home-based Business	321 TAYLOR DR	WINDSOR	5089	1
21.	21	507956.286665811	112526.647412444	HARPOON BREWERY	1.8375	Food Establishment	Restaurant	336 RUTH CARNEY DR	WINDSOR	5089	1
22.	22	509295.05546735	109206.512448289	NIDAS	0.05	Food Establishment	Home-based Business	35 N MAIN ST	WINDSOR	5089	1
23.	23	509295.05546735	109206.512448289	NIDAS	0.05	Food Establishment	Home-based Business	35 N MAIN ST	WINDSOR	5089	1

Streaming Dataset 1: Created a project as shown in the below image.

OpenRefine Food Waste in Dining Services_Streaming Permalink

Facet / Filter Undo / Redo 7 / 7

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

3480 rows

sector year sub_sector sub_sector_category food_type tons_surplus tons_supply us_dollars_surplus tons_waste tons uneaten tons_eaten tons_not_fit_for_h

All

1. 2021 Farm Not Applicable Not Applicable Dry Goods 172560.30 6469628.88 26204987.36 172477.99 172582.24 8.06131517280007 72392.173840573

2. 2021 Farm Not Applicable Not Applicable Produce 15334908.83 94986275.79 1063838160.27 13353219.21 15084480.56 250428.2697878384 6873701.35240692

3. 2021 Foodservice Bars And Taverns Bars And Taverns 722.27 27144.46 3639887.47 712.00 712.25 10.0173295262033995 311.8633987705627

4. 2021 Foodservice Bars And Taverns Bars And Taverns Dairy & Eggs 343.59 12912.94 183457.97 338.70 338.83 4.7653620783384705 75.8232970338982

5. 2021 Foodservice Bars And Taverns Bars And Taverns Dry Goods 614.17 23081.81 2120144.54 605.43 605.65 8.51809673410775 159.89898508658305

6. 2021 Foodservice Bars And Taverns Bars And Taverns Fresh Meat & Seafood 1463.77 55011.28 7844208.19 1442.94 1443.47 20.30123824963853 799.1316529646197

7. 2021 Foodservice Bars And Taverns Bars And Taverns Frozen 3.30 124.33 9148.55 3.26 3.26 0.04588566901541233 0.0

8. 2021 Foodservice Bars And Taverns Bars And Taverns Prepared Foods 23401.33 0.0 253560061.42 23330.42 23332.01 133347.1479324143 0.0

9. 2021 Foodservice Bars And Taverns Bars And Taverns Produce 1165.38 43797.32 4157481.08 1148.80 1149.22 16.1628645965005053 1006.2584595021355

10. 2021 Foodservice Bars And Taverns Bars And Taverns Ready-To-Drink Beverages 501.26 18838.31 459528.39 494.13 494.30 6.952049725962712 34.1164384058852

11. 2021 Foodservice Business & Industry Business & Industry Breads & Bakery 2080.79 120678.84 7984162.31 2047.80 2048.79 32.004899332374 1466.7406149926028

12. 2021 Foodservice Business & Industry Business & Industry Dairy & Eggs 3108.94 179411.82 10565445.09 3059.66 3061.13 47.81247209964121 746.30073417196682

13. 2021 Foodservice Business & Industry Business & Industry Dry Goods 6836.60 394528.17 20259914.45 6728.23 6731.46 105.14004373071738 2538.443937688394

14. 2021 Foodservice Business & Industry Business & Industry Fresh Meat & Seafood 5316.41 306801.14 39805545.88 5232.14 5234.65 81.76117261244163 3712.8488542515283

15. 2021 Foodservice Business & Industry Business & Industry Frozen 314.12 18127.39 868602.53 309.14 309.29 4.8308724635995335 0.0

16. 2021 Foodservice Business & Industry Business & Industry Prepared Foods 219078.15 0.0 153979866.28 218509.70 218535.01 118818.4042699213 0.0

17. 2021 Foodservice Business & Industry Business & Industry Produce 4323.44 249498.23 13653427.85 4254.91 4256.95 66.4901965334916 3811.822310165822

18. 2021 Foodservice Business & Industry Business & Industry Ready-To-Drink Beverages 2825.09 163031.21 2371969.71 2780.31 2781.64 43.447111485785724 0.0

19. 2021 Foodservicer Caters Caters Breads & Bakery 1244.01 46752.17 4773368.99 1223.50 1224.15 19.859767115531092 537.1368738336861

20. 2021 Foodservicer Caters Caters Dairy & Eggs 1858.69 69853.20 6316601.05 1828.06 1829.02 29.67281378729601 410.16965913495727

21. 2021 Foodservicer Caters Caters Dry Goods 4087.29 153607.81 12112486.35 4019.93 4022.04 65.25077545260608 1064.11624872454

22. 2021 Foodservicer Caters Caters Fresh Meat & Seafood 3178.44 119451.67 23797932.85 3126.06 3127.70 50.74165585408887 1735.2370384091114

23. 2021 Foodservicer Caters Caters Frozen 187.79 7057.82 519298.10 184.70 184.80 2.998079163538019 0.0

24. 2021 Foodservicer Caters Caters Prepared 209592.54 0.0 2303739935.25 207385.04 207395.37 335113.80855779565 0.0

Transforming the column data to store up to 3 digits after decimal in all the required columns.

OpenRefine Food Waste in Dining Services_Streaming Permalink

Facet / Filter Undo / Redo 6 / 6

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

3480 rows

Using facets and filters

Custom text transform on column tons uneaten

Expression Language General Refine Expression Language (GREL) No syntax error.

substring(value.toString(), 0, indexOf(value.toString(), '.') + 3)

Preview History Starred Help

On error keep original Re-transform up to 10 times until no change set to blank store error

OK Cancel

rs_surplus tons_waste tons uneaten tons_eaten tons not fit for h

1. 2021 Farm business & industry Prepared Foods 219078.10 0.0 130919866.28 218509.70 218535.01 118818.4042699213 0.0

2. 2021 Farm business & industry Produce 4323.44 249498.23 13653427.85 4254.91 4256.95 66.4901965334916 3811.822310165822

3. 2021 Foodservice business & industry Ready-To-Drink Beverages 2825.09 163031.21 2371969.71 2780.31 2781.64 43.447111485785724 0.0

4. 2021 Foodservice Caters Caters Breads & Bakery 1244.01 46752.17 4773368.99 1223.50 1224.15 19.859767115531092 537.1368738336861

5. 2021 Foodservice Caters Caters Dairy & Eggs 1858.69 69853.20 6316601.05 1828.06 1829.02 29.67281378729601 410.16965913495727

6. 2021 Foodservice Caters Caters Dry Goods 4087.29 153607.81 12112486.35 4019.93 4022.04 65.25077545260608 1064.11624872454

7. 2021 Foodservice Caters Caters Fresh Meat & Seafood 3178.44 119451.67 23797932.85 3126.06 3127.70 50.74165585408887 1735.2370384091114

8. 2021 Foodservice Caters Caters Frozen 187.79 7057.82 519298.10 184.70 184.80 2.998079163538019 0.0

9. 2021 Foodservice Caters Caters Prepared 209592.54 0.0 2303739935.25 207385.04 207395.37 335113.80855779565 0.0

Transforming columns from text to number format and creating a cluster for Food_type, sector, sub_sector, cause_group and cause_name and combining all the similar names into one.

OpenRefine Food Waste in Dining Services_Streaming

Facet / Filter Undo / Redo 16 / 16

13140 rows

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

Extensions Wikibase

1. 2021 Farm Not Applicable Facet Rejections Buyer Rejections 12,998 10945.162 0
2. 2021 Farm Not Applicable Text filter Packhouse Losses (Not Inedible) 10,039 8453.373 0
3. 2021 Farm Not Applicable Edit cells Transform... 35318.157 18915865.536 0 35318.157
4. 2021 Farm Not Applicable Edit column Common transforms Fill down 0 0 0
5. 2021 Farm Not Applicable Transpose Blank down 799.762 514722.057 0
6. 2021 Farm Not Applicable Sort... View Split multi-valued cells... 581.689 313881.741 0 581.689
7. 2021 Farm Not Applicable Reconcile Join multi-valued cells... Cluster and edit... 1830.39 865972.673 0 1830.3
8. 2021 Farm Not Applicable Dry Goods Not Harvested Harvested (Pests/Disease) 34529.555 98338493.095 0 34528.55
9. 2021 Farm Not Applicable Dry Goods Not Harvested Left Behind After Harvest (Inedible) 58236.533 77236591.175 0
10. 2021 Farm Not Applicable Dry Goods Not Harvested Left Behind After Harvest (Marketable) 41138.799 65747753.438 0
11. 2021 Farm Not Applicable Dry Goods Spoiled Packhouse Losses (Inedible) 133.38 112309.106 0 133.3
12. 2021 Farm Not Applicable Produce Buyer Rejections Buyer Rejections 57909.688 224831268.93 136236.146
13. 2021 Farm Not Applicable Produce Excess Packhouse Losses (Not Marketable) 317456.205 150294394.454 68438.934
14. 2021 Farm Not Applicable Produce Not Harvested Fields Never Harvested (Bad Weather) 461460.963 170951117.159 124697.424 461460.96
15. 2021 Farm Not Applicable Produce Not Harvested Fields Never Harvested (Food Safety) 52.318 17378.596 22.778 52.31
16. 2021 Farm Not Applicable Produce Not Harvested Fields Never Harvested (Market Dynamics) 746.79 289411.247 264.951
17. 2021 Farm Not Applicable Produce Not Harvested Fields Never Harvested (Other) 27537.05 9893980.173 5920.981 27537.05

OpenRefine ReFED US Food Surplus Cause Summary csv

Facet / Filter Undo / Redo 8 / 8

13140 rows

Find groups of different cell values that might be other representations of the same thing. For example, "New York" and "new york" likely refer to the same concept and just differ by capitalization, and "Gödel" and "Godel" probably refer to the same person. Find out more...

Method: Key collision Keying function: Daitch-Mokotoff

Cluster and edit column "cause_name"

3 clusters found

Cluster size	Row count	Values in cluster	Merge?	New cell value	# Choices in cluster	# Rows in cluster	Average length of choices	Length variance of choices
5	120	<ul style="list-style-type: none"> Fields Never Harvested (Bad Weather) (24 rows) Fields Never Harvested (Food Safety) (24 rows) Fields Never Harvested (Market Dynamics) (24 rows) Fields Never Harvested (Other) (24 rows) Fields Never Harvested (Pests/Disease) (24 rows) 	<input type="checkbox"/>	Fields Never Harvested (Bad Weather)	2 — 5	48 — 120	30 — 38.67	2.494 — 3.347
3	72	<ul style="list-style-type: none"> Left Behind After Harvest (Inedible) (24 rows) Left Behind After Harvest (Marketable) (24 rows) Left Behind After Harvest (Not Marketable) (24 rows) 	<input type="checkbox"/>	Left Behind After Harvest (Inedible)	2 — 5	48 — 120	30 — 38.67	2.494 — 3.347
2	48	<ul style="list-style-type: none"> Packhouse Losses (Inedible) (24 rows) Packhouse Losses (Not Marketable) (24 rows) 	<input type="checkbox"/>	Packhouse Losses (Inedible)	2 — 5	48 — 120	30 — 38.67	2.494 — 3.347

Selected all Deselect all Export clusters Merge selected & re-cluster Merge selected & Close Close

1. 2021 Farm Not Applicable Produce Not Harvested Safety) Fields Never Harvested (Market Dynamics) 746.79 289411.247 264.951
2. 2021 Farm Not Applicable Produce Not Harvested Fields Never Harvested (Other) 27537.05 9893980.173 5920.981 27537.05

Streaming Dataset 2: Created a project as shown in the below image.

OpenRefine Food Waste in Dining Services_Streaming_2 Permalink

Facet / Filter Undo / Redo 0 / 0

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

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

date_collection_datetime lbs_collected compost_created_lbs

	date_collection_datetime	lbs_collected	compost_created_lbs
1.	2022-02-09	59.0	6.0
2.	2022-02-14	329.4	33.0
3.	2022-03-21	601.2	60.0
4.	2022-04-29	419.8	42.0
5.	2022-05-16	724.8	72.0
6.	2022-05-27	535.4	54.0
7.	2022-06-03	672.8	67.0
8.	2022-06-22	511.8	51.0
9.	2022-06-27	680.8	68.0
10.	2022-06-29	574.8	57.0
11.	2022-07-15	426.8	43.0
12.	2022-07-18	807.8	81.0
13.	2022-07-29	318.8	32.0
14.	2022-08-03	531.8	53.0
15.	2022-08-19	699.8	70.0
16.	2022-09-09	709.8	71.0
17.	2022-09-14	464.8	46.0
18.	2022-09-26	562.4	56.0
19.	2022-09-30	556.6	56.0
20.	2022-10-07	454.8	45.0
21.	2022-10-10	653.4	65.0
22.	2022-10-31	953.4	95.0
23.	2022-11-18	713.6	71.0
24.	2022-11-28	844.4	84.0
25.	2022-12-05	834.4	83.0
26.	2023-01-03	730.4	73.0
27.	2023-01-06	968.6	97.0
28.	2023-01-20	752.4	75.0
29.	2023-01-30	642.4	64.0
30.	2023-02-10	571.6	57.0
31.	2023-02-15	536.8	54.0
32.	2023-03-10	670.6	67.0
33.	2023-03-13	779.4	78.0

Changed the column to date format.

OpenRefine Food Waste in Dining Services_Streaming_2 Permalink Text transform on 267 cells in column date_collection_datetime: value.toDate() Undo

Facet / Filter Undo / Redo 1 / 1

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

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

date_collection_datetime lbs_collected compost_created_lbs

	date_collection_datetime	lbs_collected	compost_created_lbs
1.	2022-02-09T00:00:00Z	59.0	6.0
2.	2022-02-14T00:00:00Z	329.4	33.0
3.	2022-03-21T00:00:00Z	601.2	60.0
4.	2022-04-29T00:00:00Z	419.8	42.0
5.	2022-05-16T00:00:00Z	724.8	72.0
6.	2022-05-27T00:00:00Z	535.4	54.0
7.	2022-06-03T00:00:00Z	672.8	67.0
8.	2022-06-22T00:00:00Z	511.8	51.0
9.	2022-06-27T00:00:00Z	680.8	68.0
10.	2022-09-09T00:00:00Z	574.8	57.0
11.	2022-07-15T00:00:00Z	426.8	43.0
12.	2022-07-18T00:00:00Z	807.8	81.0
13.	2022-07-29T00:00:00Z	318.8	32.0
14.	2022-08-03T00:00:00Z	531.8	53.0
15.	2022-08-19T00:00:00Z	699.8	70.0
16.	2022-09-09T00:00:00Z	709.8	71.0
17.	2022-09-14T00:00:00Z	464.8	46.0
18.	2022-09-26T00:00:00Z	562.4	56.0
19.	2022-09-30T00:00:00Z	556.6	56.0
20.	2022-10-07T00:00:00Z	454.8	45.0
21.	2022-10-10T00:00:00Z	653.4	65.0
22.	2022-10-31T00:00:00Z	953.4	95.0
23.	2022-11-18T00:00:00Z	713.6	71.0
24.	2022-11-28T00:00:00Z	844.4	84.0
25.	2022-12-05T00:00:00Z	834.4	83.0
26.	2023-01-03T00:00:00Z	730.4	73.0
27.	2023-01-06T00:00:00Z	968.6	97.0
28.	2023-01-20T00:00:00Z	752.4	75.0
29.	2023-01-30T00:00:00Z	642.4	64.0
30.	2023-02-10T00:00:00Z	571.6	57.0
31.	2023-02-15T00:00:00Z	536.8	54.0
32.	2023-03-10T00:00:00Z	670.6	67.0
33.	2023-03-13T00:00:00Z	779.4	78.0

Changing the format of columns from text to numeric format and verified for null values in the dataset.

OpenRefine Food Waste in Dining Services_Streaming_2

267 rows

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

Facet / Filter Undo / Redo 3/3

Refresh Reset all Remove all

date_collection_datetime lbs_collected compost_created_lbs

All 1. 2022-02-09T00:00:00Z 59 6
2. 2022-02-14T00:00:00Z 329.4 33
3. 2022-03-21T00:00:00Z 601.2 60
4. 2022-04-29T00:00:00Z 419.8 42
5. 2022-05-16T00:00:00Z 724.8 72
6. 2022-05-27T00:00:00Z 535.4 54
7. 2022-06-03T00:00:00Z 672.8 67
8. 2022-06-22T00:00:00Z 511.8 51
9. 2022-06-27T00:00:00Z 680.8 68
10. 2022-06-29T00:00:00Z 574.8 57
11. 2022-07-15T00:00:00Z 426.8 43
12. 2022-07-18T00:00:00Z 807.8 81
13. 2022-07-29T00:00:00Z 316.8 32
14. 2022-08-03T00:00:00Z 531.8 53
15. 2022-08-19T00:00:00Z 699.8 70
16. 2022-09-09T00:00:00Z 709.8 71
17. 2022-09-14T00:00:00Z 464.8 46
18. 2022-09-26T00:00:00Z 582.4 56
19. 2022-09-30T00:00:00Z 556.6 56
20. 2022-10-07T00:00:00Z 454.8 45
21. 2022-10-10T00:00:00Z 653.4 65
22. 2022-10-31T00:00:00Z 953.4 95
23. 2022-11-18T00:00:00Z 713.6 71
24. 2022-11-28T00:00:00Z 844.4 84
25. 2022-12-05T00:00:00Z 834.4 83
26. 2023-01-03T00:00:00Z 730.4 73
27. 2023-01-06T00:00:00Z 966.6 97
28. 2023-01-20T00:00:00Z 752.4 75
29. 2023-01-30T00:00:00Z 642.4 64
30. 2023-02-10T00:00:00Z 571.8 57
31. 2023-02-15T00:00:00Z 536.8 54
32. 2023-03-10T00:00:00Z 670.6 67
33. 2023-03-13T00:00:00Z 779.4 78

Data Storage and Management:

Created a project named “ADTA5240F23G15” in GCP.

Dashboard - ADTA5240F23G15

Start your free trial with \$300 in credit. Don't worry – you won't be charged if you run out of credit. [Learn more](#) DISMISS START FREE

Google Cloud ADTA5240F23G15

DASHBOARD ACTIVITY RECOMMENDATIONS CUSTOMISE

Project info

- Project name: ADTA5240F23G15
- Project number: 151487202459
- Project ID: adta5240f23g15

[ADD PEOPLE TO THIS PROJECT](#)

[Go to project settings](#)

Resources

- BigQuery Data warehouse/analytics
- SQL Managed MySQL, PostgreSQL, SQL Server
- Compute Engine VMs, GPUs, TPUs, disks
- Storage Multi-class multi-region object storage
- Cloud Functions Event-driven serverless functions

API APIs

Requests (requests/sec)

No data is available for the selected time frame.

[Go to APIs overview](#)

Google Cloud Platform status

Google Cloud SQL

Some multiple-table DMLs intermittently are not working on MySQL versions 8.0.30+ with query insights enabled.

Began at 2023-11-27 (05:01:07)

All times are US/Pacific

Data provided by status.cloud.google.com

[Go to Cloud status dashboard](#)

Billing

Estimated charges USD \$0.00 For the billing period 1–29 Nov 2023

[Take a tour of billing](#)

[View detailed charges](#)

Monitoring

Create my dashboard

Now viewing project ADTA5240F23G15' in organisation 'No organisation'

Created a storage bucket and stored files in it.

The screenshot shows the Google Cloud Storage interface for the bucket 'g15_foodwaste'. The bucket details page is displayed, showing the location as 'us-south1 (Dallas)', storage class as 'Standard', public access as 'Not public', and protection as 'None'. The 'OBJECTS' tab is selected, showing three uploaded files: 'Food-Waste-in-Dining-Services_St...' (771.6 KB, text/csv, created 29 Nov 2023, 15:14:13), 'Food-Waste-in-Dining-Services_St...' (1.3 MB, text/csv, created 29 Nov 2023, 15:14:14), and 'Food-Waste-in-Dining-Services_St...' (7.9 KB, text/csv, created 29 Nov 2023, 15:14:13). A success message at the bottom indicates '3 files successfully uploaded'.

Created cluster along with manager and worker nodes.

The screenshot shows the Google Cloud Dataproc Clusters page. It displays a single cluster named 'g15-foodwaste' which is 'Running' in the 'us-central1' region with 2 total worker nodes. The cluster status is shown as 'Sched Off'. A message at the bottom says 'Request to create cluster g15-foodwaste submitted'.

The screenshot shows the Google Cloud Compute Engine interface. On the left, a sidebar navigation includes 'Compute Engine' (selected), 'Virtual machines' (selected), 'Instance templates', 'Sole-tenant nodes', 'Machine images', 'TPUs', 'Committed-use discounts', 'Reservations', and 'Migrate to Virtual Machine...'. Under 'Storage', there are 'Disks', 'Snapshots', 'Images', 'Marketplace', and 'Release notes'. The main content area displays 'VM instances' with tabs for 'INSTANCES', 'OBSERVABILITY', and 'INSTANCE SCHEDULES'. Three VM instances are listed:

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
Running	g15-foodwaste-m	us-central1-a			10.128.0.3	35.224.244.160	SSH
Running	g15-foodwaste-w-0	us-central1-a			10.128.0.2	35.193.111.160	SSH
Running	g15-foodwaste-w-1	us-central1-a			10.128.0.4	34.123.228.160	SSH

Below the table, 'Related actions' include 'Explore Backup and DR', 'View billing report', 'Monitor VMs', 'Explore VM logs', 'Set up firewall rules', and 'Patch management'. A modal window titled 'Request to create cluster g15-foodwaste submitted' is open. On the right, a sidebar titled 'Get started with Compute Engine' provides links to tutorials for creating websites on IIS and Apache, and for transferring files to Windows and Linux VMs. It also includes a link to 'Configure firewall rules'.

BigQuery: Enabling BigQuery API and executed queries for data verification for both static and streaming datasets. (Few screenshots attached here)

The screenshot shows the Google Cloud BigQuery Studio interface. The left sidebar has 'Explorer' (selected), 'SCHEMATIC' (selected), and 'SUMMARY' (selected). The main area shows a query editor with the code:

```
1 SELECT Type,sum(TonsPerWeek) as Sum_TonsPerWeek FROM `adta5240f23g15.dataset1.static` GROUP BY type;
```

The 'RESULTS' tab is selected, displaying the following data:

Row	Type	Sum_TonsPerWeek
1	Food Establishment	2039.135906177...
2	Food Manufacturer	394.7630388578...
3	Hospital	3.58249999999...
4	Nursing/Assisted Living	34.1092500000...
5	School	5.657955
6	Correctional Facility	8.78325

BigQuery - ADTA5240F23G11

console.cloud.google.com/bigquery?project=adta5240f23g15&ws=!1m5!1m4!m3!1sadta5240f23g15!2sdataset1!3ssstatic!1m0

Start your free trial with \$300 in credit. Don't worry – you won't be charged if you run out of credit. [Learn more](#)

DISMISS START FREE

Google Cloud ADTA5240F23G15 Search (/) for resources, docs, products and more

Explorer + ADD I < > + Q Untitled 2 RUN SAVE DOWNLOAD SHARE SCHEDULE MORE DISMISS

Type to search

Found 0 results.

adt5240f23g15 External connections dataset1 static

SCHEMA DETAILS

Filter Enter property

Field name

ID Latitude Longitude FSGName TonsPerWeek TYPE Sub_Type Address County ZIP Active

RESULTS PREVIEW

JOB INFORMATION

Query results

Sum_TonsPerWeek

Row	Sub_Type	Sum_TonsPerWeek
1	Bakery	85.59
2	School	34.22
3	Caterer	71.08
4	Lodging	183.06
5	Restaurant	1375.41
6	Camp	27.54
7	Mobile Unit	11.37
8	Grocery	124.89
9	Food Processor	13.5
10	Convenience Store	43.56
11	Home-based Business	12.83
12	Feed/Animal Products	4.04
13	Bar	19.51
14	Sweets	2.41

RESULTS CHART PREVIEW

JSON EXECUTION DETAILS EXECUTION GRAPH

Press Option+F1 for accessibility options.

SUMMARY

Nothing currently selected

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM and creating Python notebooks with Colab.

EDIT SCHEMA VIEW ROW

Job history

REFRESH

Results per page: 50 1 – 31 of 31

Welcome – Google Cloud

BigQuery - ADTA5240F23G11

console.cloud.google.com/bigquery?authuser=1&project=adta5240f23g15&ws=!1m5!1m4!m3!1sadta5240f23g15!2sg15streaming!13ssstreaming!1m0

Start your free trial with \$300 in credit. Don't worry – you won't be charged if you run out of credit. [Learn more](#)

DISMISS START FREE

Google Cloud ADTA5240F23G15 Search (/) for resources, docs, products and more

BigQuery Explorer + ADD I < > + Q Untitled 2 RUN SAVE RESULTS EXPLORE DATA

Type to search

Viewing resources. SHOW STARRED ONLY

BigQuery Studio

Data transfers Scheduled queries Analytics Hub Dataform Partner Centre

Migration Assessment SQL translation

Administration Monitoring Capacity management BI Engine Release notes

adt5240f23g15 dataset1 g15streaming1 streaming1

SCHEMA DETAILS

Filter Enter property name or

Field name

year sector sub_sector food_type cause_group cause_name tons_surplus_due_to cause_us_dollars tons_inedible_parts tons_not_fit_for_human_cause

RESULTS PREVIEW

JOB INFORMATION

Query results

Sum_usdolars_Surpl

Row	sector	Sum_usdolars_Surpl	Sum_tons_Surplus
1	Foodservice	1620496708103.0	142792396.0
2	Residential	2266991075597.0	477928223.0
3	Retail	267867236023.0	53634789.0
4	Farm	122411488956.0	199927161.0
5	Manufacturing	397533834249.0	147095529.0

RESULTS CHART PREVIEW

JSON EXECUTION DETAILS EXECUTION GRAPH

Press Option+F1 for accessibility options.

SUMMARY

Nothing currently selected

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM and creating Python notebooks with Colab.

EDIT SCHEMA VIEW ROW

Job history

REFRESH

The screenshot shows the Google Cloud BigQuery interface. On the left, the navigation pane includes sections for Analysis, Migration, Administration, and BigQuery Studio. In the center, the Explorer pane shows datasets like 'dataset1' and 'g15foodwastestreaming2'. The main area displays a query titled 'Untitled' with the following SQL code:

```

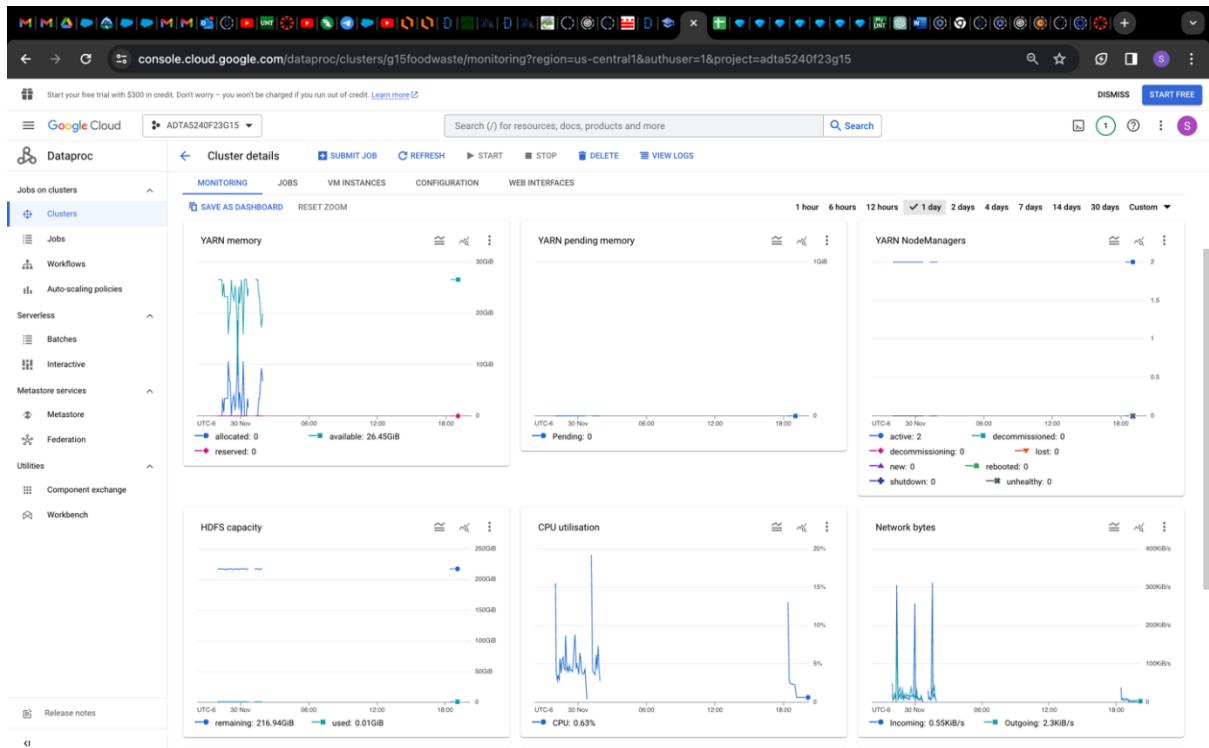
7 SELECT * FROM `adta5240f23g15.g15foodwastestreaming2.streaming2`
8 ORDER BY Food_Waste_Collected DESC
9 LIMIT 10;

```

The 'Query results' section shows a table with 10 rows of data:

Row	Collection_Date	Food_Waste_Collected	Estimated_Earned_C
1	2023-10-02 00:00:00 UTC	1316.0	132
2	2023-09-05 00:00:00 UTC	1257.4	126
3	2023-07-31 00:00:00 UTC	1227.4	123
4	2023-09-11 00:00:00 UTC	1109.4	111
5	2023-07-10 00:00:00 UTC	1104.4	110
6	2023-04-17 00:00:00 UTC	1098.4	110
7	2023-07-03 00:00:00 UTC	1080.4	108
8	2022-08-10 00:00:00 UTC	1065.8	107
9	2023-06-12 00:00:00 UTC	1050.4	105
10	2023-06-05 00:00:00 UTC	1016.4	102

Further monitored the results of the cluster and monitored various parameters from it.



Data Analysis:

This is performed using Hive and Spark SQL executing various queries for both static and streaming datasets. For this created a project directory and copied dataset files to the

directory using Hadoop Linux commands.

```
Linux g15foodwaste-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2-bpo10+1 (2022-07-28) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
sharathreddy1240@g15foodwaste:~$ hdfs dfs -ls /
Found 3 items
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:57 /tmp
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /var
sharathreddy1240@g15foodwaste:~$ hdfs dfs -ls /user
Found 11 items
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/dataproc
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/dbbase
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/hbase
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/hive
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/homedir
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/hql
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/solr
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/spark
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/storm
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/zeppelin
drwxrwxrwt - hdfs hadoop 0 2023-11-30 03:56 /user/zookeeper
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data/foodwaste_static
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data/foodwaste_streaming_1
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~$ hdfs dfs -ls /user/g15foodwaste/data
Found 0 items
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_static
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_streaming_1
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~$ pwd
/home/sharathreddy1240
sharathreddy1240@g15foodwaste:~$ cd DATA
sharathreddy1240@g15foodwaste:~/DATA$ ls -l
total 0
sharathreddy1240@g15foodwaste:~/DATA$
```

```
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data/foodwaste_streaming_1
sharathreddy1240@g15foodwaste:~$ hdfs dfs -mkdir /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~$ hdfs dfs -ls /user/g15foodwaste/data
Found 0 items
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_static
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_streaming_1
drwxr-xr-x - sharathreddy1240 hadoop 0 2023-11-30 04:03 /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~$ pwd
/home/sharathreddy1240
sharathreddy1240@g15foodwaste:~$ ls -l
total 0
sharathreddy1240@g15foodwaste:~$ mkdir DATA
sharathreddy1240@g15foodwaste:~$ cd DATA
sharathreddy1240@g15foodwaste:~/DATA$ ls -l
total 0
sharathreddy1240@g15foodwaste:~/DATA$ gsutil cp gs://g15foodwaste/data/foodwaste_static.csv foodwaste_static.csv
Copying gs://g15foodwaste/data/foodwaste_static.csv...
/ [1 files] [771.6 KiB/771.6 KiB]
Operation completed over 1 objects/771.6 KiB.
sharathreddy1240@g15foodwaste:~/DATA$ gsutil cp gs://g15foodwaste/data/Foodwaste_streaming_1.csv foodwaste_streaming_1.csv
Copying gs://g15foodwaste/data/Foodwaste_streaming_1.csv...
/ [1 files] [1.3 MiB/ 1.3 MiB]
Operation completed over 1 objects/1.3 MiB.
sharathreddy1240@g15foodwaste:~/DATA$ gsutil cp gs://g15foodwaste/data/Foodwaste_streaming_2.csv foodwaste_streaming_2.csv
Copying gs://g15foodwaste/data/Foodwaste_streaming_2.csv...
/ [1 files] [ 7.9 KiB/ 7.9 KiB]
Operation completed over 1 objects/7.9 KiB.
sharathreddy1240@g15foodwaste:~/DATA$ ls -l
total 2120
-rw-r--r-- 1 sharathreddy1240 sharathreddy1240 790125 Nov 30 04:08 foodwaste_static.csv
-rw-r--r-- 1 sharathreddy1240 sharathreddy1240 1370688 Nov 30 04:08 foodwaste_streaming_1.csv
-rw-r--r-- 1 sharathreddy1240 sharathreddy1240 8065 Nov 30 04:09 foodwaste_streaming_2.csv
sharathreddy1240@g15foodwaste:~/DATA$ gsutil cp gs://g15foodwaste/data/foodwaste_static gs://g15foodwaste/data/foodwaste_static
sharathreddy1240@g15foodwaste:~/DATA$ hdfs dfs -put foodwaste_streaming_1.csv /user/g15foodwaste/data/foodwaste_streaming_1
sharathreddy1240@g15foodwaste:~/DATA$ hdfs dfs -put foodwaste_streaming_2.csv /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~/DATA$ hdfs dfs -ls /user/g15foodwaste/data/foodwaste_static
Found 1 items
-rw-r--r-- 2 sharathreddy1240 sharathreddy1240 790125 2023-11-30 04:10 /user/g15foodwaste/data/foodwaste_static/foodwaste_static.csv
sharathreddy1240@g15foodwaste:~/DATA$ hdfs dfs -ls /user/g15foodwaste/data/foodwaste_streaming_1
Found 1 items
-rw-r--r-- 2 sharathreddy1240 sharathreddy1240 1370686 2023-11-30 04:11 /user/g15foodwaste/data/foodwaste_streaming_1/foodwaste_streaming_1.csv
sharathreddy1240@g15foodwaste:~/DATA$ hdfs dfs -ls /user/g15foodwaste/data/foodwaste_streaming_2
sharathreddy1240@g15foodwaste:~/DATA$ pwd
/home/sharathreddy1240/DATA
sharathreddy1240@g15foodwaste:~/DATA$ cd
sharathreddy1240@g15foodwaste:~$ pwd
/home/sharathreddy1240
sharathreddy1240@g15foodwaste:~$
```

Hive:

Static Dataset:

Now, connecting to the **hive** and creating tables for the dataset to generate results for some queries.

```
Welcome - Google Cloud console | g15_foodwaste_2.csv - Ob | ssh.cloud.google.com/v2/ssl |
SSH-in-browser
+-----+
| tab_name |
+-----+
+-----+
No rows selected (1.637 seconds)
0: jdbc:hive2://localhost:10000> create external table if not exist foodwaste-static
(foodwaste-static string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/';
Error: Error while compiling statement: FAILED: ParseException line 1:29 missing KW EXISTS at 'exist' near 'exist' in create table statement
line 1:35 cannot recognize input near 'foodwaste' '-' 'static' in create table statement (state=42000,code=40000)
0: jdbc:hive2://localhost:10000> create external table if not exists foodwaste-static
(foodwaste-static string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/';
Error: Error while compiling statement: FAILED: ParseException line 1:45 cannot recognize input near 'foodwaste' '-' 'static' in table name (state=42000,code=40000)
0: jdbc:hive2://localhost:10000> create external table if not exists foodwastestatic
(fooodwaste-static string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/';
Error: Error while compiling statement: FAILED: ParseException line 2:10 cannot recognize input near '-' 'static' 'string' in column type (state=42000,code=40000)
0: jdbc:hive2://localhost:10000> create external table if not exists foodwastestatic
(foodwastestatic string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/';
INFO : Compiling command(queryId=hive_20231130002124_d49dbd67-a481-4cfe-a84b-63d9da33a38e): create external table if not exists foodwastestatic
(foodwastestatic string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/'
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Analysis Completed (read 0 rows)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20231130002124_d49dbd67-a481-4cfe-a84b-63d9da33a38e); Time taken: 0.209 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130002124_d49dbd67-a481-4cfe-a84b-63d9da33a38e): create external table if not exists foodwastestatic
(foodwastestatic string)
row format delimited
stored as textfile
location '/user/g15foodwaste/data/foodwaste-static/'
INFO : Starting task [Stage@0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20231130002124_d49dbd67-a481-4cfe-a84b-63d9da33a38e); Time taken: 1.646 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (2.776 seconds)
0: jdbc:hive2://localhost:10000>
```

Query1: SELECT TYPE, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, County
FROM foodwastestatic_1
Group by TYPE, County
Order by County;:

	type	sum_tonsperweek	county
Food Establishment	10.13	BRADFORD	
Food Manufacturer	0.0	BRADFORD	
Nursing/Assisted Living	0.48	BRADFORD	
Food Establishment	0.18	BRAINTREE	
Food Manufacturer	0.0	BRAINTREE	
Food Establishment	15.78	BRIGHTON	
Food Manufacturer	0.0	BRANDON	
Nursing/Assisted Living	0.04	BRANDON	
Food Establishment	81.84	BRATTLEBORO	
Food Manufacturer	0.13	BRATTLEBORO	
Nursing/Assisted Living	1.09	BRATTLEBORO	
School	0.3	BRATTLEBORO	
School	0.02	BRIDGEWATER	
Food Establishment	5.35	BRIDGEWATER	
Food Establishment	0.61	BURKE	
Food Establishment	2.75	BRIGHTON	
Food Manufacturer	0.0	BRIGHTON	
Food Establishment	8.29	BRISTOL	
Food Manufacturer	0.0	BRISTOL	
Nursing/Assisted Living	0.08	BRISTOL	
Food Establishment	1.39	BROOKFIELD	
Food Manufacturer	0.0	BROOKFIELD	
School	0.02	BROOKLINE	
Food Establishment	0.05	BROOKLINE	
Food Establishment	0.31	BROWNSINGTON	
Food Establishment	0.06	BROWNSVILLE	
Food Establishment	8.3	BURKE	
Food Manufacturer	0.0	BURKE	
Food Establishment	149.29	BURLINGTON	
Food Manufacturer	6.7	BURLINGTON	
Nursing/Assisted Living	2.36	BURLINGTON	
School	0.46	BURLINGTON	
Food Establishment	1.11	CABOT	
Food Manufacturer	0.0	CABOT	
Food Establishment	1.98	CALAIS	
Food Manufacturer	0.02	CALAIS	
Food Establishment	20.84	CAMBRIDGE	
Food Manufacturer	0.0	CAMBRIDGE	
Food Establishment	3.45	CANAAN	
Food Manufacturer	0.0	CANAAN	
Food Establishment	15.92	CASTLETON	
Nursing/Assisted Living	0.13	CASTLETON	
Food Establishment	4.47	CAVENDISH	
Food Establishment	0.42	CHARLESTON	
Food Establishment	4.86	CHARLOTTE	
Food Manufacturer	0.0	CHARLOTTE	
Food Establishment	1.33	CHELSEA	

Query 2: SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1 Group by county;

```
0: jdbc:hive2://localhost:10000> SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1
INFO : Compiling command(queryId=hive_20231130044220_190ab755-3ab4-4523-b631-0bcb421b0336): SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1
Group by county
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:sum_tonsperweek, type:double, comment:null), FieldSchema(name:county, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20231130044220_190ab755-3ab4-4523-b631-0bcb421b0336); Time taken: 0.209 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130044220_190ab755-3ab4-4523-b631-0bcb421b0336): SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1
Group by county
INFO : Query ID = hive_20231130044220_190ab755-3ab4-4523-b631-0bcb421b0336
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20231130044220_190ab755-3ab4-4523-b631-0bcb421b0336
INFO : Session is already open
INFO : Dag name: SELECT round(sum(TonsPerWeek),2) as... .county (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0003)

----- VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED -----
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
-----
```

SSH-in-browser

```

| 1.38 | ADDISON
| 0.68 | ALBANY
| 4.55 | ALBURGH
| 4.91 | ARLINGTON
| 1.18 | ASCUTNEY
| 0.92 | AVERILL
| 1.37 | BAKERSFIELD
| 1.99 | BARNARD
| 0.49 | BARNET
| 33.54 | BART
| 4.83 | BARRE CITY
| 10.28 | BARRE TOWN
| 5.91 | BARTON
| 0.0 | BEEBE PLAIN
| 17.39 | BELLOWS FALLS
| 48.85 | BENNINGTON
| 0.84 | BENSON
| 2.08 | BERKSHIRE
| 14.29 | BERNIN
| 5.2 | BETHEL
| 0.26 | BLOOMFIELD
| 8.1 | BOLTON
| 0.43 | BOMOSEEN
| 4.73 | BONDVILLE
| 10.61 | BRADFORD
| 0.18 | BRAINTREE
| 15.82 | BRANDON
| 83.37 | BRADFORDSBORO
| 5.57 | BRIDGEWATER
| 0.61 | BRIDPORT
| 2.75 | BRIGHTON
| 8.37 | BRISTOL
| 1.39 | BROOKFIELD
| 0.07 | BROOKLINE
| 0.31 | BROWNINGTON
| 0.06 | BROWNSVILLE
| 8.3 | BURKE
| 158.62 | BURLINGTON
| 1.11 | CABOT
| 2.0 | CALAIS
| 20.84 | CAMBRIDGE
| 3.45 | CANAN
| 14.05 | CASTLETON
| 4.47 | CAVENDISH
| 0.42 | CHARLESTON
| 4.86 | CHARLOTTE
| 14.4 | CHELSEA
| 17.48 | CHESTER
+-----+
| sum_tonsperweek | county |
+-----+

```

sum_tonsperweek county

Query 3: SELECT * FROM foodwastestatic_1
WHERE type='Food Establishment' AND Sub_Type='Restaurant';

```

0: jdbc:hive2://localhost:10000> SELECT * FROM foodwastestatic_1
+-----+
| _id | type | sub_type | longitude | latitude | address | feaname | county | zip | tonsperweek | active |
+-----+
| 508517.092486237 | Restaurant | 1 | 110239.699568204 | JUNIPER HILL INN | WINDSOR | 0509 | 0.735 | 1 | Food Establishment
| 509053.226834364 | Restaurant | 1 | 108528.483279731 | WINDSOR ELKS BLDGE 1665 | WINDSOR | 0509 | 1.68 | 1 | Food Establishment
| 508685.55860496 | Restaurant | 1 | 111316.571731314 | JOHN P LARKIN COUNTRY CLUB | WINDSOR | 0509 | 0.945 | 1 | Food Establishment
| 509050.434552743 | Restaurant | 1 | 110606.020730204 | STUBS & LAURAS | WINDSOR | 0509 | 0.26 | 1 | Food Establishment
| 509050.434552743 | Restaurant | 1 | 110606.020730204 | FRAZERS PLACE | WINDSOR | 0509 | 0.42 | 1 | Food Establishment
| 507891.405948801 | Restaurant | 1 | 112693.219421738 | AMERICAN CRAFTED SPIRITS INCORPORATED | WINDSOR | 0509 | 0.2625 | 1 | Food Establishment
| 507956.286665811 | Restaurant | 1 | 112526.647412444 | HARPOON BREWERY | WINDSOR | 0509 | 1.8375 | 1 | Food Establishment
| 509157.934644762 | Restaurant | 1 | 109873.029570039 | BOSTON DREAMS LLC | WINDSOR | 0509 | 0.105 | 1 | Food Establishment
| 509190.234211642 | Restaurant | 1 | 108917.512897908 | CLARAS CAFE | WINDSOR | 0509 | 0.26 | 1 | Food Establishment
| 509190.234211642 | Restaurant | 1 | 108917.512897908 | CROW BAR & GRILL | WINDSOR | 0509 | 0.26 | 1 | Food Establishment
| 507899.647931704 | Restaurant | 1 | 112559.910357247 | SUSTAINABLE FARMER THIS | WINDSOR | 0509 | 0.525 | 1 | Food Establishment
| 509270.522519229 | Restaurant | 1 | 109362.01649555 | YUMMY CHINESE RESTAURANT | WINDSOR | 0509 | 0.441 | 1 | Food Establishment
| 509157.979955755 | Restaurant | 1 | 108839.700129574 | PIZZA CHEF OF WINDSOR | WINDSOR | 0509 | 0.966 | 1 | Food Establishment
| 495684.505512832 | Restaurant | 1 | 116022.949693854 | SOUTH WOODSTOCK COUNTRY STORE | WOODSTOCK | 0509 | 0.1155 | 1 | Food Establishment
| 498474.596845076 | Restaurant | 1 | 124776.429715926 | RED ROOSTER | WOODSTOCK | 0509 | 0.26 | 1 | Food Establishment
| 498490.803722531 | Restaurant | 1 | 125065.288519136 | PRINCE AND THE PAINTER | WOODSTOCK | 0509 | 0.735 | 1 | Food Establishment
| 24 ELM ST | Restaurant | 1 | 125065.288519136 | PRINCE AND THE PAINTER | WOODSTOCK | 0509 | 1 | Food Establishment
+-----+

```

Query 4: select count(type) as Count_Type, zip, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1

```
where type='Food Establishment'  
group by zip;
```

Query 5: select Sub_Type, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1 group by Sub_Type;

```

0: jdbc:hive2://localhost:10000> select Sub_Type,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
group by Sub_Type
INFO : Compiling command(queryId:hive_20231130045048_7990b33f-dfd9-4493-8fad-1e55368d6c71): select Sub_Type,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
group by Sub_Type
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retained = false)
INFO : Returning Hive schema: Schema([FieldSchemas:[FieldSchema(name:sub_type, type:string, comment:null), FieldSchema(name:sum_tonsperweek, type:double, comment:null)], properties:map[]])
INFO : Executing compiled command(queryId:hive_20231130045048_7990b33f-dfd9-4493-8fad-1e55368d6c71): select Sub_Type,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
group by Sub_Type
INFO : Query ID = hive_20231130045048_7990b33f-dfd9-4493-8fad-1e55368d6c71
INFO : Total jobs = 1
INFO :Launching Job 1 out of 1
INFO : Running with mapred.job.tracker=192.168.1.10:19920
INFO : Submitted to counters: {} for queryId: hive_20231130045048_7990b33f-dfd9-4493-8fad-1e55368d6c71
INFO : Session is already open
INFO : Dag name: select Sub_Type,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
INFO : Status: Running (Executing on YARN cluster with App Id application_1701316582184_0003)

-----  

VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map Reducer 1 container INITED 1 0 0 0 1 0 0
Reducer 2 container INITED 1 0 0 0 1 0 0
-----  

VERTICES: 0/0 [(>-----)] 0% ELAPSED TIME: 2.10 s

```

```

INFO : Completed executing command(queryId=hive_20231130045048_7990b33f-dfd9-4493-8fad-1e55368d6c71); Time taken: 6.518 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+-----+
| sub_type | sum_tonsperweek |
+-----+-----+
| 0.378 | 0.0 |
| Alcohol Manufacturer | 0.02 |
| B&B | 25.09 |
| Bakery | 85.33 |
| Bar | 13.51 |
| Cafeteria | 8.78 |
| Camp | 27.54 |
| Caterer | 71.03 |
| Church | 0.06 |
| Coffee Roaster | 0.0 |
| College/University | 7.74 |
| Concession Stand | 7.36 |
| Convenience Store | 41.56 |
| Dairy | 356.83 |
| Distribution | 0.0 |
| Feed/Animal Products | 4.04 |
| Food Establishment | 0.0 |
| Food Manufacturer | 0.0 |
| Food Processor | 13.5 |
| Grocery | 124.71 |
| Homebased Business | 13.33 |
| Hospital | 13.43 |
| Juice/Cider | 0.13 |
| Lodging | 183.06 |
| Map products | 0.0 |
| Meat/Poultry/Seafood | 0.18 |
| Mobile Unit | 11.37 |
| Nursing/Assisted Living | 0.0 |
| Pet Shop | 1.86 |
| Residential Facility | 32.12 |
| Restaurant | 1365.78 |
| School | 34.22 |
| Self-service meals | 1.1 |
| Sub_Type | 0.0 |
| Sweets | 2.41 |
| Vending | 0.05 |
+-----+-----+
36 rows selected (6.73 seconds)
0: jdbc:hive2://localhost:10000>

```

Streaming Dataset 1:

Query 1: SELECT cause_group, year, round (sum (tons_surplus_due_to_cause),2)as Sum_tons_Surplus FROM foodwaste_streaming_1
group by year, cause_group
order by year, Sum_tons_Surplus;

```

0: jdbc:hive2://localhost:10000> SELECT cause_group, year, round (sum ( tons_surplus_due_to_cause),2)as Sum_tons_Surplus FROM foodwaste_streaming_1
... > group by year, cause_group order by year, Sum_tons_Surplus;
INFO : Compiling command(queryId=hive_20231130052127_eab520ec-f69c-46ac-8638-b09ec8020263): SELECT cause_group, year, round (sum ( tons_surplus_due_to_cause),2)as Sum_tons_Surplus
FROM foodwaste_streaming_1
group by year, cause_group order by year, Sum_tons_Surplus
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:cause_group, type:string, comment:null), FieldSchema(name:year, type:string, comment:null), FieldSchema(name:sum_tons_surplus, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20231130052127_eab520ec-f69c-46ac-8638-b09ec8020263); Time taken: 0.236 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130052127_eab520ec-f69c-46ac-8638-b09ec8020263): SELECT cause_group, year, round (sum ( tons_surplus_due_to_cause),2)as Sum_tons_Surplus
FROM foodwaste_streaming_1
group by year, cause_group order by year, Sum_tons_Surplus
INFO : Queued for hive_20231130052127_eab520ec-f69c-46ac-8638-b09ec8020263
INFO : Total jobs: 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20231130052127_eab520ec-f69c-46ac-8638-b09ec8020263
INFO : Session is already open
INFO : Dag name: SELECT cause_group, year,...Sum_tons_Surplus (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0005)

-----+-----+-----+-----+-----+-----+-----+-----+
    VERTICES      MODE      STATUS    TOTAL    COMPLETED    RUNNING    PENDING    FAILED    KILLED
-----+-----+-----+-----+-----+-----+-----+-----+
Map 1        container    INITED      0       0       0       1       0       0
Reducer 2    container    INITED      1       0       0       1       0       0
Reducer 3    container    INITED      1       0       0       1       0       0
-----+-----+-----+-----+-----+-----+-----+-----+
VERTICES: 00/03  [>>-----] 0%   ELAPSED TIME: 1.53 s

```

INFO : Concurrency mode is disabled, not creating a lock manager			
cause_group	year	sum_tons	surplus
Mistakes & Malfunctions	2010	1300957.19	
Buyer Rejections	2010	1374725.52	
Other	2010	2046673.04	
Food Safety	2010	2353700.25	
Date Label Concerns	2010	4980534.43	
Spoiled	2010	1.188039336E7	
Not Harvested	2010	1.383565892E7	
Excess	2010	1.714149478E7	
Trimmings & Byproducts	2010	2.121474582E7	
Mistakes & Malfunctions	2010	1332320.42	
Buyer Rejections	2011	1466430.0	
Other	2011	140000.05	
Food Safety	2011	2431656.22	
Date Label Concerns	2011	5186495.87	
Spoiled	2011	1.225297478E7	
Not Harvested	2011	1.3807325E7	
Excess	2011	1.766593516E7	
Trimmings & Byproducts	2011	2.239316353E7	
Mistakes & Malfunctions	2012	1361592.54	
Buyer Rejections	2012	1505052.09	
Other	2012	2158959.93	
Food Safety	2012	2489770.63	
Date Label Concerns	2012	5340794.96	
Spoiled	2012	1.239609193E7	
Not Harvested	2012	1.390929554E7	
Excess	2012	1.767878181E7	
Trimmings & Byproducts	2012	2.269716539E7	
Mistakes & Malfunctions	2013	1397451.73	
Buyer Rejections	2013	1547293.52	
Other	2013	2236783.43	
Food Safety	2013	2583865.96	
Date Label Concerns	2013	5538620.92	
Spoiled	2013	1.271871274E7	
Not Harvested	2013	1.394535889E7	
Excess	2013	1.887620529E7	
Trimmings & Byproducts	2013	2.343475891E7	
Mistakes & Malfunctions	2014	1428201.47	
Buyer Rejections	2014	1851927.7	
Other	2014	2268816.12	
Food Safety	2014	2618240.65	
Date Label Concerns	2014	570011.1	
Spoiled	2014	1.3003160687	
Not Harvested	2014	1.411423249E7	
Excess	2014	1.937960749E7	
Trimmings & Byproducts	2014	2.421360089E7	
Mistakes & Malfunctions	2015	1454430.3	

INFO : Concurrency mode is disabled, not creating a lock manager			
cause_group	year	sum_tons	surplus
Other	2017	2336438.7	
Food Safety	2017	2555361.85	
Date Label Concerns	2017	6158906.0	
Spoiled	2017	1.300186713E7	
Not Harvested	2017	1.493783579E7	
Excess	2017	2.0390404462E7	
Trimmings & Byproducts	2017	2.426260998E7	
Mistakes & Malfunctions	2018	1488229.4	
Buyer Rejections	2018	1.581622.15	
Other	2018	2460917.37	
Food Safety	2018	2733045.79	
Date Label Concerns	2018	6571231.81	
Spoiled	2018	1.396419584E7	
Not Harvested	2018	2.305146853E7	
Excess	2018	2.525651769E7	
Trimmings & Byproducts	2018	2.565506948E7	
Mistakes & Malfunctions	2019	1.594635.88	
Buyer Rejections	2019	1.594635.88	
Other	2019	2423637.7	
Food Safety	2019	2755419.52	
Date Label Concerns	2019	6607050.61	
Spoiled	2019	1.389348044E7	
Not Harvested	2019	2.368214892E7	
Excess	2019	2.565506948E7	
Trimmings & Byproducts	2019	2.565506948E7	
Buyer Rejections	2020	1590676.33	
Mistakes & Malfunctions	2020	1651396.36	
Other	2020	2652567.04	
Food Safety	2020	3082402.12	
Date Label Concerns	2020	6623399.22	
Spoiled	2020	1.394535889E7	
Not Harvested	2020	428892235E7	
Excess	2020	1.942853985E7	
Trimmings & Byproducts	2020	2.744542789E7	
Mistakes & Malfunctions	2021	1450434.46	
cause_group	year	sum_tons	surplus
Buyer Rejections	2021	1616169.35	
Other	2021	2557772.14	
Food Safety	2021	2958992.54	
Date Label Concerns	2021	7007652.93	
Not Harvested	2021	1.272039109E7	
Spoiled	2021	1.383422158E7	
Excess	2021	2.156001158E7	
Trimmings & Byproducts	2021	2.726553455E7	
cause_group	year	0.0	

```
Query 2: SELECT sub_sector, round (sum(us_dollars_surplus_due_to_cause)) as  
Sum_usdollars_Surplus, round(sum( tons_surplus_due_to_cause) ) as Sum_tons_Surplus  
FROM foodwaste_streaming_1  
group by sub_sector;
```

```
Welcome - Google Cloud console BigQuery - ADTA5240F23G1 ssh.cloud.google.com/v2/ssh/projects/adta5240f23g1/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024... Downloads Hive Analysis: Setup & Query + SSH-in-browser UPLOAD FILE DOWNLOAD FILE ... : INFO : Semantic Analysis Completed (retryal = false)
INFO : Returning Hive schema: Schema(fieldsSchema:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20231130052512_c500dd2e-73c2-4796-a804-86049a81fc5a); Time taken: 0.026 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130052512_c500dd2e-73c2-4796-a804-86049a81fc5a): show tables
INFO : Starting task [Stage=0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20231130052512_c500dd2e-73c2-4796-a804-86049a81fc5a); Time taken: 0.014 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| tab_name |
+-----+
| foodwaste_streaming_1 |
| foodwastesstatic_1 |
+-----+
2 rows selected (0.208 seconds)
0: jdbc:hive2://localhost:10000> SELECT sub_sector, round(sum(us_dollars_surplus_due_to_cause)) as
. . . . . > Sum_usdollars_Surplus, round(sum( tons_surplus_due_to_cause) ) as Sum_tons_Surplus
. . . . . > FROM foodwaste_streaming_1
. . . . . group by sub_sector
. . . . .
Error: Error while compiling statement: FAILED: SemanticException [Error 10004]: Line 4:9 Invalid table alias or column reference 'sub_sector': (possible column names are: year, sector, sub-sector, food_type, cause_group, cause_name, tons_surplus_due_to_cause, us_dollars_surplus_due_to_cause, tons_inedible_parts, tons_not_fit_for_human_consumption) (state=400,code=10004)
0: jdbc:hive2://localhost:10000> select sub-sector, round(sum(us dollars surplus due to cause)) as Sum USDollars Surplus,
. . . . . > round(sum(tons surplus due to cause) ) as Sum tons Surplus from foodwaste_streaming_1
. . . . . > group by sub-sector;
Error: Error while compiling statement: FAILED: SemanticException [Error 10004]: Line 3:9 Invalid table alias or column reference 'sub': (possible column names are: year, sector, sub-sector, food_type, cause_group, cause_name, tons_surplus_due_to_cause, us_dollars_surplus_due_to_cause, tons_inedible_parts, tons_not_fit_for_human_consumption) (state=42000,code=10004)
0: jdbc:hive2://localhost:10000> ALTER TABLE your_table CHANGE old_column new_column STRING;
Error: Error while compiling statement: FAILED: SemanticException [Error 10001]: Table not found under the state (state=42S02,code=10001)
0: jdbc:hive2://localhost:10000> alter table foodwaste_streaming_1 change `sub-sector` `sub_sector` string;
INFO : Compiling command(queryId=hive_20231130053134_e21b49cf-9b27-4678-bb77-dd8ff7b9009b); alter table foodwaste_streaming_1 change `sub-sector` `sub_sector` string
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retryal = false)
INFO : Returning Hive schema: Schema(fieldsSchema:[], properties:null)
INFO : Completed compiling command(queryId=hive_20231130053134_e21b49cf-9b27-4678-bb77-dd8ff7b9009b); Time taken: 0.064 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130053134_e21b49cf-9b27-4678-bb77-dd8ff7b9009b): alter table foodwaste_streaming_1 change `sub-sector` `sub_sector` string
INFO : Starting task [Stage=0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20231130053134_e21b49cf-9b27-4678-bb77-dd8ff7b9009b); Time taken: 0.181 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
No rows affected (0.257 seconds)
0: jdbc:hive2://localhost:10000> SELECT sub_sector, round(sum(us_dollars_surplus_due_to_cause)) as
. . . . . > Sum_usdollars_Surplus, round(sum( tons_surplus_due_to_cause) ) as Sum_tons_Surplus
. . . . . > FROM foodwaste_streaming_1
. . . . . group by sub_sector;
```

```
Welcome - Google Cloud console | BigQuery - ADTA5240F23C1 | ssh.cloud.google.com/v2/ssh | Downloads | Hive Analysis: Setup & Query +  
← → G % ssh.cloud.google.com/v2/ssh/projects/adta5240f23g15/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024... ... Open Star Download File Help More  
SSH-in-browser UPLOAD FILE DOWNLOAD FILE ! Help More  
  
INFO : Completed compiling command(queryId=hive_20231130053249_2ad6263d-0e31-4bd6-9744-a1384b51d33b); Time taken: 0.162 seconds  
INFO : Concurrency mode is disabled, not creating a lock manager  
INFO : Executing command(queryId=hive_20231130053249_2ad6263d-0e31-4bd6-9744-a1384b51d33b): SELECT sub_sector, round (sum(us_dollars_surplus_due_to_cause)) as sum_tons_surplus FROM foodwaste_streaming_1 group by sub_sector  
INFO : Query ID = hive_20231130053249_2ad6263d-0e31-4bd6-9744-a1384b51d33b  
INFO : Total jobs = 1  
INFO : Launching Job 1 out of 1  
INFO : Starting execution of job 1 in serial mode  
INFO : Submits to counters: [] for queryId:hive_20231130053249_2ad6263d-0e31-4bd6-9744-a1384b51d33b  
INFO : Tex session hasn't been created yet. Opening session  
INFO : Dag name: SELECT sub_sector, round (sum(...).sub_sector (Stage-1)  
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0006)  
  
INFO : Completed executing command(queryId=hive_20231130053249_2ad6263d-0e31-4bd6-9744-a1384b51d33b); Time taken: 15.169 seconds  
INFO : OK  
INFO : Concurrency mode is disabled, not creating a lock manager  
  


| VERTICES  | MODE      | STATUS    | TOTAL | COMPLETED | RUNNING | PENDING | FAILED | KILLED |
|-----------|-----------|-----------|-------|-----------|---------|---------|--------|--------|
| Map 1     | container | SUCCEEDED | 1     | 1         | 0       | 0       | 0      | 0      |
| Reducer 2 | container | SUCCEEDED | 1     | 1         | 0       | 0       | 0      | 0      |

  
VERTICES: 02/02 [Completed] (0 rows) ->| 100% ELAPSED TIME: 7.89 s  
  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
| sub_sector | sum_usdollars_surplus | sum_tons_surplus |  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
| Bars And Taverns | 3.719313407E9 | 375396.0 |  
| Business & Industry | 3.477200048E10 | 350033.0 |  
| Cafeteria | 2.535380048E10 | 310063.0 |  
| Corrections | 7.496702038E9 | 1068095.0 |  
| Full Service Restaurants | 8.80854303794E11 | 6.612368E7 |  
| Healthcare | 6.285570522E10 | 3002153.0 |  
| K-12 Education | 5.399115077E10 | 7512744.0 |  
| Limited Service Restaurants | 1.494524614E11 | 1494524614.0 |  
| Lodging | 1.494524614E11 | 1494524614.0 |  
| Military | 7.339233106E9 | 1139537.0 |  
| Not Applicable | 3.05480363482E12 | 1054331.0 |  
| Other | 4.99636422E9 | 688127.0 |  
| Recreation | 7.2933734659E10 | 6812361.0 |  
| Refreshment Services | 1.9750726E10 | 4153573.0 |  
| Transportation | 1.4368304923E10 | 1510925.0 |  
| sub_sector | 0.0 | 0.0 |  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
17 rows selected (17.103 seconds)  
0: jdbc:hive2://localhost:10000  
0: jdbc:hive2://localhost:10000>
```

```
Query 3: SELECT sector, round (sum(us_dollars_surplus_due_to_cause)) as  
Sum_usdollars_Surplus, round(sum(tons_surplus_due_to_cause) )as Sum_tons_Surplus  
FROM foodwaste_streaming_1  
group by sector;
```

Query 4: SELECT sector, cause_group FROM foodwaste_streaming_1 group by sector, cause_group order by sector;

```

0: jdbc:hive2://localhost:10000> SELECT sector, cause_group FROM foodwaste_streaming_1 group by sector, cause_group order by sector, cause_group;
INFO : Compiling command(queryId=hive_20231130053619_c9031628-44c3-451d-bef6-ee0cafe0bd3c): SELECT sector, cause_group FROM foodwaste_streaming_1 group by sector, cause_group order by sector, cause_group
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:sector, type:string, comment:null), FieldSchema(name:cause_group, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20231130053619_c9031628-44c3-451d-bef6-ee0cafe0bd3c); Time taken: 0.21 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130053619_c9031628-44c3-451d-bef6-ee0cafe0bd3c): SELECT sector, cause_group FROM foodwaste_streaming_1 group by sector, cause_group order by sector, cause_group
r by sector
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20231130053619_c9031628-44c3-451d-bef6-ee0cafe0bd3c
INFO : Session is already open
INFO : Dag name: SELECT sector, cause_group FROM foo...sector (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0006)

-----  

  VERTICES      MODE      STATUS  TOTAL COMPLETED RUNNING PENDING FAILED KILLED  

Map 1    container    INITED   1      0      0      1      0      0  

Reducer 2  container  INITED   1      0      0      1      0      0  

Reducer 3  container  INITED   1      0      0      1      0      0  

-----  

VERTICES: 0/0  [>>-----] 0%  ELAPSED TIME: 1.15 s
-----  

|-----+-----+  

|   sector |   cause_group |  

|-----+-----+  

| Farm    | Buyer Rejections |  

| Farm    | Not Harvested |  

| Farm    | Excess |  

| Farm    | Spoiled |  

| Foodservice | Mistakes & Malfunctions |  

| Foodservice | Date Label Concerns |  

| Foodservice | Excess |  

| Foodservice | Other |  

| Foodservice | Spoiled |  

| Foodservice | Trimmings & Byproducts |  

| Manufacturing | Buyer Rejections |  

| Manufacturing | Excess |  

| Manufacturing | Trimmings & Byproducts |  

| Residential | Excess |  

| Residential | Date Label Concerns |  

| Residential | Food Safety |  

| Residential | Mistakes & Malfunctions |  

| Residential | Other |  

| Residential | Spoiled |  

| Residential | Trimmings & Byproducts |  

| Retail    | Date Label Concerns |  

| Retail    | Excess |  

| Retail    | Food Safety |  

| Retail    | Mistakes & Malfunctions |  

| Retail    | Other |  

| Retail    | Spoiled |  

| Retail    | Trimmings & Byproducts |  

| sector    | cause_group |  

|-----+-----+  

28 rows selected (6.842 seconds)
0: jdbc:hive2://localhost:10000>

```

Query 5: `SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1 where year=2021 group by food_type order by Sum_tons_Surplus limit 1;`

```

0: jdbc:hive2://localhost:10000> SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1
0: jdbc:hive2://localhost:10000> where year=2021
0: jdbc:hive2://localhost:10000> group by food_type
0: jdbc:hive2://localhost:10000> order by Sum_tons_Surplus limit 1;
INFO : Compiling command(queryId=hive_20231130053755_d0c5d442-2973-4b39-97af-d9a6d9bfff8f3): SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1
where year=2021
group by food_type
order by Sum_tons_Surplus limit 1
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:food_type, type:string, comment:null), FieldSchema(name:sum_tons_surplus, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20231130053755_d0c5d442-2973-4b39-97af-d9a6d9bfff8f3); Time taken: 0.167 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130053755_d0c5d442-2973-4b39-97af-d9a6d9bfff8f3): SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1
where year=2021
group by food_type
order by Sum_tons_Surplus limit 1
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20231130053755_d0c5d442-2973-4b39-97af-d9a6d9bfff8f3
INFO : Session is already open
INFO : Dag name: SELECT food_type, round (sum (tons surpl... (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0006)

-----  

  VERTICES      MODE      STATUS  TOTAL COMPLETED RUNNING PENDING FAILED KILLED  

Map 1    container    INITED   0      0      1      0      0  

Reducer 2  container  INITED   1      0      0      1      0      0  

Reducer 3  container  INITED   1      0      0      1      0      0  

-----  

VERTICES: 0/0  [>>-----] 0%  ELAPSED TIME: 2.25 s
-----  


```

```

INFO : Completed executing command(queryId=hive_20231130053755_d0c5d442-2973-4b39-97af-d9a6d9bff8f3); Time taken: 6.707 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| food_type | sum_tons_surplus |
+-----+
| Breads & Bakery | 2151409.52 |
+-----+
1 row selected (6.905 seconds)
0: jdbc:hive2://localhost:10000> Closing: 0: jdbc:hive2://localhost:10000

```

Streaming Dataset 2:

```

Welcome - Google Cloud console | BigQuery - ADTA5240F23G | ssh.cloud.google.com/v2/ssh/projects/adta5240f23g15/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024... | Downloads | Hive Analysis: Setup & Query | +
SSH-in-browser
INFO : Estimated Earned Compost Created: string
...
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| food_type | sum_tons_surplus |
+-----+
| Breads & Bakery | 2151409.52 |
+-----+
1 row selected (6.905 seconds)
0: jdbc:hive2://localhost:10000> Closing: 0: jdbc:hive2://localhost:10000

```

Query 1: SELECT * FROM foodwaste_streaming_2
order by Food_Waste_Collected desc
limit 10;

```

Welcome - Google Cloud console | BigQuery - ADTA5240F23G | ssh.cloud.google.com/v2/ssh/projects/adta5240f23g15/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024... | Downloads | Hive Analysis: Setup & Query | +
SSH-in-browser
limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Session is already completed (status: false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:foodwaste_streaming_2.collection_date, type:string, comment:null), FieldSchema(name:foodwaste_streaming_2.food_waste_collected, type:string, comment:null), FieldSchema(name:foodwaste_streaming_2.estimated_earned_compost_created, type:string, comment:null)]), properties:null
INFO : Completed compiling command(queryId=hive_20231130060544_da52f4e1-637d-4b5a-b954-261be3af0d45); Time taken: 0.163 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231130060544_da52f4e1-637d-4b5a-b954-261be3af0d45): SELECT * FROM foodwaste_streaming_2
...
INFO : Total jobs: 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Specified to run counter: [] for queryId: hive_20231130060544_da52f4e1-637d-4b5a-b954-261be3af0d45
INFO : Tez session hasn't been created yet. Opening session
INFO : Dag name: select * from foodwaste_streaming_2 limit 10 (Stage-1)
INFO : Status: Running (Reusing on YARN cluster with App id application_1701316582184_0009)
INFO : Completed executing command(queryId=hive_20231130060544_da52f4e1-637d-4b5a-b954-261be3af0d45); Time taken: 14.829 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 0 0 0 0 0 0
+-----+
VERTICES 01/01 [=====>>>] 100% ELAPSED TIME: 7.96 s
+-----+
| foodwaste_streaming_2.collection_date | foodwaste_streaming_2.food_waste_collected | foodwaste_streaming_2.estimated_earned_compost_created |
| Collection Date | Food Waste Collected | Estimated Earned Compost Created |
| 2022-02-09T00:00:00Z | 59 | 6 |
| 2022-03-10T00:00:00Z | 109.4 | 109 |
| 2022-03-11T00:00:00Z | 601.2 | 601 |
| 2022-04-29T00:00:00Z | 419.8 | 419.8 |
+-----+
5 rows selected (17.077 seconds)
0: jdbc:hive2://localhost:10000>

```

Query 2: SELECT * FROM foodwaste_streaming_2
where Collection_Date like '2023-06-19%';

```

Welcome - Google Cloud com | BigQuery - ADTA5240F23G15 | ssh.cloud.google.com/v2/ssh | Downloads | Hive Analysis: Setup & Query ...
ssh.cloud.google.com/v2/ssh/projects/adta5240f23g15/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024...
SSH-in-browser
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0 0
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 5.05 s
INFO : Completed executing command(queryId:hive_20231130060711_1196629f-b6d1-4ade-958a-25b474c56784); Time taken: 5.767 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| foodwaste_streaming_2.collection_date | foodwaste_streaming_2.food_waste_collected | foodwaste_streaming_2.estimated_earned_compost_created |
+-----+
+-----+
No rows selected (5.94 seconds)
0: jdbc:hive2://localhost:10000> SELECT * FROM foodwaste_streaming_2
   WHERE Collection_Date like '2023-06-19';
INFO : Compiling command(queryId:hive_20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e): SELECT * FROM foodwaste_streaming_2
where Collection Date like '2023-06-19'
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retry= false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:foodwaste_streaming_2.collection_date, type:string, comment:null), FieldSchema(name:foodwaste_streaming_2.food_waste_collected, type:string, comment:null), FieldSchema(name:foodwaste_streaming_2.estimated_earned_compost_created, type:string, comment:null)]), properties:null
INFO : Completed compiling command(queryId:hive_20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e); Time taken: 0.189 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId:hive_20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e): SELECT * FROM foodwaste_streaming_2
where Collection Date like '2023-06-19'
INFO : Query ID = 20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e
INFO : Total jobs = 1
INFO : Launching Job 1 out of 1
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Subscribed to counters: [] for queryId: hive_20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e
INFO : Session is already open
INFO : Data name: SELECT * FROM foodwaste_stre...'2023-06-19%' (Stage-1)
INFO : Status: Running (Executing on YARN cluster with App id application_1701316582184_0010)

VERTICES MODE STATUS TOTAL_COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 01/01 [=====>>>] 100% ELAPSED TIME: 5.42 s
INFO : Completed executing command(queryId:hive_20231130060736_2420a167-e54e-467f-8cb1-f1db3710f82e); Time taken: 6.084 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager
+-----+
| foodwaste_streaming_2.collection_date | foodwaste_streaming_2.food_waste_collected | foodwaste_streaming_2.estimated_earned_compost_created |
+-----+
| 2023-06-19T00:00:00Z | 0 | 0 |
+-----+
1 row selected (6.302 seconds)
0: jdbc:hive2://localhost:10000>

```

Spark:

Static Dataset:

Now, connecting to the **spark sql** and generated results for some queries.

Query 1: `SELECT TYPE, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, County
FROM foodwastestatic_1
Group by TYPE, County
Order by County;`

SSH-in-browser

```

Food Establishment 0.05 ADAMANT
Food Establishment 1.38 ADDISON
Food Manufacturer 0.1 ALBANY
Food Establishment 0.68 ALBANY
Food Establishment 4.55 ALCBURGH
Food Manufacturer 0.0 ALCBURGH
Food Manufacturer 0.0 ARLINGTON
Food Establishment 4.91 ARLINGTON
Food Establishment 1.1 ASCUTNEY
Food Establishment 0.08 ASCUTNEY
Food Establishment 0.92 AVERILL
Food Establishment 1.37 BAKERSFIELD
Food Establishment 1.95 BARNARD
School 0.04 BARNARD
Food Manufacturer 0.0 BARNARD
Food Establishment 0.9 BARNET
Nursing/Assisted Living 0.03 BARTET
Food Establishment 0.05 BARRE
Nursing/Assisted Living 2.55 BARRE
Food Manufacturer 0.0 BARRE
Hospital 0.64 BARRE
Food Establishment 3.63 BARRE CITY
Food Manufacturer 1.21 BARRE CITY
Food Establishment 6.11 BARRE TOWN
Food Manufacturer 4.17 BARRE TOWN
Food Manufacturer 0.0 BARTON
Nursing/Assisted Living 0.37 BARTON
Food Establishment 5.54 BARTON
Food Manufacturer 0.0 BEEBE PLAIN
Food Establishment 17.16 BEELINE FALLS
Nursing/Assisted Living 0.23 BEELINE FALLS
Food Manufacturer 0.0 BENNINGTON
Nursing/Assisted Living 3.79 BENNINGTON
Food Establishment 44.46 BENNINGTON
Food Manufacturer 0.0 BENNINGTON
School 0.6 BENNINGTON
Food Manufacturer 0.0 BENSON
Food Establishment 0.84 BENSON
Food Establishment 2.08 BERKSHIRE
Food Establishment 13.26 BERLIN
Hospital 0.77 BERLIN
Food Manufacturer 0.26 BERLIN
School 0.12 BETHEL
Food Establishment 5.08 BETHEL
Food Establishment 0.22 BLOOMFIELD
Food Establishment 8.1 BOLTON
Food Establishment 0.39 BOMOESEN
Nursing/Assisted Living 0.03 BOMOESEN
Food Establishment 4.73 BONDVILLE
Nursing/Assisted Living 0.48 BRADFORD

```

SSH-in-browser

```

Nursing/Assisted Living 0.04 WESTFORD
Food Manufacturer 0.18 WESTMINSTER
School 0.04 WESTMINSTER
Nursing/Assisted Living 0.04 WESTMINSTER
Food Establishment 2.34 WESTMINSTER
Food Establishment 1.67 WESTMORE
Food Establishment 4.17 WESTON
Food Establishment 0.12 WEYBRIDGE
Food Establishment 0.11 WHEELOCK
Nursing/Assisted Living 0.67 WHITE RIVER JUNCTION
Food Establishment 13.65 WHITE RIVER JUNCTION
Hospital 0.32 WHITE RIVER JUNCTION
Food Manufacturer 0.0 WHITE RIVER JUNCTION
Food Establishment 0.18 WHITING
School 0.03 WHITING
Food Establishment 0.87 WHITINGHAM
Food Manufacturer 0.0 WHITINGHAM
School 0.06 WHITINGHAM
Food Establishment 0.05 WILDER
Nursing/Assisted Living 0.12 WILLIAMSTOWN
Food Manufacturer 0.0 WILLIAMSTOWN
Food Establishment 3.5 WILLIAMSTOWN
Food Establishment 21.05 WILLISTON
School 0.05 WILLISTON
Food Manufacturer 0.78 WILLISTON
Food Establishment 22.04 WILMINGTON
School 0.03 WILMINGTTON
School 0.01 WINDHAM
Food Establishment 0.84 WINDHAM
Food Establishment 13.22 WINDSOR
Correctional Facility 0.53 WINDSOR
Hospital 0.18 WINDSOR
Food Manufacturer 0.0 WINDSOR
Nursing/Assisted Living 0.8 WINDSOR
Food Establishment 4.62 WINHALL
Food Manufacturer 0.05 WINOOSKI
School 0.0 WINOOSKI
Nursing/Assisted Living 0.23 WINOOSKI
Food Establishment 12.24 WINOOSKI
Food Establishment 0.44 WOLCOTT
Food Establishment 0.78 WOODBURY
Food Establishment 1.26 WOODFORD
Food Manufacturer 0.6 WOODSTOCK
School 0.12 WOODSTOCK
Nursing/Assisted Living 0.46 WOODSTOCK
Food Establishment 14.65 WOODSTOCK
Food Establishment 0.42 WORCESTER
Nursing/Assisted Living 0.06 WORCESTER
Time taken: 17.583 seconds, Fetched 610 row(s)
spark-sql> 
```

Query 2: SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1 Group by county;

SSH-in-browser

```

Food Establishment 4.62 WINHALL
Food Manufacturer 0.05 WINOOSKI
School 0.0 WINOOSKI
Nursing/Assisted Living 0.23 WINOOSKI
Food Establishment 12.24 WINOOSKI
Food Establishment 0.44 WOLCOTT
Food Establishment 0.78 WOODBURY
Food Establishment 1.26 WOODFORD
Food Manufacturer 0.6 WOODSTOCK
School 0.12 WOODSTOCK
Nursing/Assisted Living 0.46 WOODSTOCK
Food Establishment 14.65 WOODSTOCK
Food Establishment 0.42 WORCESTER
Nursing/Assisted Living 0.06 WORCESTER
Time taken: 17.583 seconds, Fetched 610 row(s)
spark-sql> > SELECT round(sum(TonsPerWeek),2) as Sum_TonsPerWeek, county FROM foodwastestatic_1
> Group by county;
0.87 WEST NEWBURY
1.74 GLOVER
0.45 RT 244"
1.66 LOWELL
0.52 STRAFFORD
2.19 GREENSBORO
15.83 WOODSTOCK
0.15 SOUTH STRAFFORD
0.75 TOPSHAM
NULL County
0.11 EAST CORINTH
13.27 RANDOLPH
NULL 100 SO CRAFTSBURY RD
NULL 502 MAIN STREET
1.73 CRAFTSBURY COMMON
3.62 WILLIAMSTOWN
0.15 WASHINGTON
0.38 VERSHIRE
1.02 SOUTH WOODSTOCK
0.42 CHARLESTON
6.59 NEWBURY
NULL 747 PUTNEY RD
3.65 TOWNSHEND
83.37 BRATTLEBORO
0.81 DUMMERSTON
5.16 WEST DOVER
0.45 GUILDFORD
1.91 NEWPORT CITY
70.94 DOVER
1.98 GRAFTON
98.9 RUTLAND
NULL 1 GABLES PLACE

```

SSH-in-browser

```

0.1 EAST TOPSHAM
1.37 CHITTENDEN
0.23 HUBBARDTON
1.24 MIDDLETON SPRINGS
0.13 IRA
0.92 WARDSBORO
8.56 FAIR HAVEN
NULL 1169 SPRING LAKE ROAD
0.13 EAST FOULNEY
38.25 STRATTON
48.85 BENNINGTON
0.04 WEST WARDSBORO
22.07 WILMINGTON
0.05 WILDER
NULL VOCATIONAL CENTER
0.05 WEST TOWNSHEND
5.2 BETHEL
0.05 BONDVILLE
NULL 604 US ROUTE 7 SOUTH
5.5 DORSET
0.42 LANDGROVE
10.92 PERU
NULL 147 INDUSTRIAL PARK DRIVE
NULL 19 PINE HAVEN SHORE ROAD
0.05 ADAMANT
2.0 CALAIS
1.29 WESTFORD
1.0 ST GEORGE
0.7 EAST MONTPELIER
1.11 CABOT
NULL 15 STATE ST
4.0 SOUTH HERO
4.35 ALCBURGH
3.42 GRAND ISLE
0.04 WEST HARTFORD
3.23 ISLE LA MOTTE
3.89 NORTH HERO
0.0 EAST HARDWICK
0.81 Suite 205 "
0.0 ORANGE
NULL 35 N Main ST
NULL 14 GREENWOOD LN
0.0 HANOVER
0.0 EAST ARINGTON
1.21 WELLS RIVER
0.68 ALBANY
0.63 WATERFORD
1.32 GEORGIA
Time taken: 1.586 seconds, Fetched 357 row(s)
spark-sql> 
```

Query 3: SELECT * FROM foodwastestatic_1 where type='Food Establishment' AND Sub_Type='Restaurant';

```

spark-sql> SELECT * FROM foodwastestatic_1
> where type='Food Establishment' AND Sub_Type='Restaurant';
```

9330	485049.268022356	218880.687380712	RANCH CAMP	0	Food Establishment	Restaurant	311 MOUNTAIN RD	STOWE	5672	1
9331	442119.38968172	208897.373089121	RUSTIC ROOTS	0	Food Establishment	Restaurant	195 FALLS RD	SHELBOURNE	5482	1
9333	499715.540447699	77671.225524185	9 AND DINE RESTAURANT LLC	0	Food Establishment	Restaurant	12 COUNTRY CLUB RD	BELLOWS FALLS	5101	1
9357	476651.437892524	127305.620297518	CASEY'S CABOOSE	0	Food Establishment	Restaurant	1930 KILLINGTON RD	KILLINGTON	5751	1
9360	472811.425176	187391.000000086	WORKERS DRUG TOO	0	Food Establishment	Restaurant	111 MAD RIVER GREEN	WATTSFIELD	5673	1
9361	541652.324256128	257286.55900100	K.T. RAMS ON THE HILL	0	Food Establishment	Restaurant	109 CROSS ST	WATTSFIELD	5846	1
9364	461402.183652573	123202.832876954	RUFF LIFE COFFEE	0	Food Establishment	Restaurant	46 CENTER ST	RUTLAND	5701	1
9365	440263.185560189	223738.4055358	PHO THAO RESTAURANT	0	Food Establishment	Restaurant	1355 NORTH AVE	BURLINGTON	5401	1
9371	492310.054999749	229157.63158614	MORRISVILLE FOOD CO-OP	0	Food Establishment	Restaurant	258 WINOOSKI AVE	BURLINGTON	5661	1
9372	443626.76577245	220985.718859612	BUTCH & BABE'S	0	Food Establishment	Restaurant	31 MAIN ST	BRISTOL	5433	1
9375	453833.925118454	181546.548925744	HATCH 31	0	Food Establishment	Restaurant	4374 WOODSTOCK RD	WOODSTOCK	5091	1
9378	492063.015496813	122014.54478911	EAT WOODSTOCK	0	Food Establishment	Restaurant	D & L BEVERAGE AND DELI	HARDWICK	5843	1
9382	492119.24671314	224494.059456357	BLACK CAP COFFEE & BEER	0	Food Establishment	Restaurant	53 UPPER MAIN ST	MORRISTOWN	5661	1
9389	438201.30363535	75354.000000077	UNIVERSITY DINING	0	Food Establishment	Restaurant	4000 MAIN ST	MONTPELIER	5254	1
9393	514610.323613104	127891.943031619	PIZZERIA PIES	0	Food Establishment	Restaurant	5 8 MAIN ST	WHITE RIVER JUNCTION	5001	1
9394	503704.372188878	196897.021926648	PLAINFIELD HARDWARE	0	Food Establishment	Restaurant	4723 US ROUTE 2 EAST	MONTPELIER	5651	1
9396	475874.259965752	267480.225343028	FARM TO TABLE	0	Food Establishment	Restaurant	303 MAIN ST	ENOSBURG	5450	1
9399	440410.194644722	186554.775431201	GOD DAY CAFE	0	Food Establishment	Restaurant	11 MAIN ST	VERGENNES	5491	1
9403	442951.379198909	220169.272649758	COURTHOUSE CAFE	0	Food Establishment	Restaurant	32 CHERRY ST STE 199	BURLINGTON	5401	1
9409	522937.525219199	270698.566327009	THE WAREHOUSE	0	Food Establishment	Restaurant	150 MAIN ST	NEWPORT	5039	1
9413	521351.978076958	99922.0014520695	DOW HOME KITCHEN	0	Food Establishment	Restaurant	100 MAIN ST	MONTPELIER	5602	1
9412	463995.266406598	114262.589103106	HANGAR CAFE	0	Food Establishment	Restaurant	1002 AIRPORT RD CLARENCE	STOWE	5759	1
9413	474324.210404363	225301.391566882	THE FARM & BEERY	0	Food Establishment	Restaurant	20 HOUR COFFEE DR	SOUTH BURLINGTON	5572	1
9423	500334.1921494634	89847.0239572503	DOUGH Donuts	0	Food Establishment	Restaurant	262 RIVER ST	SPRINGFIELD	5156	1
9426	468077.994215538	46984.608881114	HERMITAGE CLUB HOUSE	0	Food Establishment	Restaurant	183 GATE HOUSE TRAIL	WILMINGTON	5363	1
9429	471216.546196827	221912.002345166	QUAD COMMONS CAFE	0	Food Establishment	Restaurant	49 PLACE ST	COLCHESTER	5446	1
9431	451465.892649241	221455.889489812	NEPALI KITCHEN	0	Food Establishment	Restaurant	10 RAILROAD AVE	ESSEX JUNCTION	5452	1
9440	462175.385487408	121260.98175019	OLYMPIC II	0	Food Establishment	Restaurant	258 S MAIN	RUTLAND	5701	1
9445	440806.652159435	223166.646201454	LA BOCA'S WOOD FIRED PIZZA	0	Food Establishment	Restaurant	1127 NORTH AVE STE 32	BURLINGTON	5401	1
9448	453996.591104174	257309.202708814	14TH STAR BREWING CO	0	Food Establishment	Restaurant	133 N MAIN ST UNIT 7	ST ALBANS CITY	5478	1
9451	443642.740147448	217329.535210475	MINIMO'S PIZZERIA	0	Food Establishment	Restaurant	408 SHELBECK RD	SOUTH BURLINGTON	5401	1
Time taken: 0.704 seconds, Fetched 2393 row(s)										

Query 4: select count(type) as Count_Type, zip, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
where type='Food Establishment'
group by zip;

```

spark-sql> select count(type) as Count_Type,zip,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
> where type='Food Establishment';
+-----+-----+-----+
|Count_Type|zip|Sum_TonsPerWeek|
+-----+-----+-----+
|      5071| 1.13|        0.34|
|      5079| 0.34|        29.1 |
|      5855| 29.1 |        0.43|
|      5859| 23.59|       23.59|
|      5829| 12.26|       12.26|
|      5045| 13.06|       13.06|
|      5076| 0.47|        0.47|
|      5372| 1.24|        0.44|
|      5060| 12.84|       12.84|
|      5081| 6.83|        0.57|
|      5830| 0.43|        0.43|
|      5072| 0.34|        0.34|
|      5820| 0.68|        0.68|
|      5845| 1.31|        0.53|
|      5053| 0.13|        0.13|
|      5833| 0.11|        0.11|
|      5037| 1.99|        1.99|
|      5030| 1.26|        1.26|
|      5040| 0.57|        0.57|
|      5842| 0.26|        0.26|
|      5091| 17.87|       17.87|
|      5058| 0.05|        0.05|
|      5051| 1.63|        1.63|
|      5675| 0.15|        0.15|
|      5860| 4.25|        4.25|
|      5827| 0.88|        0.88|
|      5041| 29.99|       29.99|
|      5345| 3.14|        3.14|
|      5362| 0.26|        0.26|
|      5857| 0.42|        0.42|
|      1109| 0.18|        0.18|
|      5346| 9.19|        9.19|
|      5303| 2.55|        2.55|
|      5658| 2.1 |        2.1 |
|      5101| 20.95|       20.95|
|      5735| 8.3 |        8.3 |
|      5764| 4.75|        4.75|
|      5701| 101.86|      101.86|
|      5351| 0.05|        0.05|
+-----+-----+-----+
```

Query 5: select Sub_Type, round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
group by Sub_Type;

```

Welcome - Google Cloud console | BigQuery - ADTA5240F23G1 | ssh.cloud.google.com/v2/ssh | Downloads | Hive Analysis: Setup & Query ...
← → C ssh.cloud.google.com/v2/ssh/projects/adta5240f23g1/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=151487202459&us...
SSH-in-browser
Time taken: 1.174 seconds, Fetched 306 row(s)
spark-sql> select Sub_Type,round(sum(TonsPerWeek),2) as Sum_TonsPerWeek from foodwastestatic_1
   > group by Sub_Type;
23/11/30 05:04:31 WARN org.apache.hadoop.util.concurrent.ExecutorHelper: Thread (Thread[GetFileInfo #1,5,min]) interrupted:
java.lang.InterruptedException
   at com.google.common.util.concurrent.AbstractFuture.get(AbstractFuture.java:510)
   at com.google.common.util.concurrent.FluentFuture$TrustedFuture.get(FluentFuture.java:88)
   at org.apache.hadoop.util.concurrent.ExecutorHelper.logThrowableFromAfterExecute(ExecutorHelper.java:48)
   at org.apache.hadoop.util.concurrent.HadoopThreadPoolExecutor.afterExecute(HadoopThreadPoolExecutor.java:90)
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1157)
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
   at java.lang.Thread.run(Thread.java:750)
College/University    7.74
Bakery    85.33
Bar      19.51
School   34.22
Sub_Type    NULL
Grocery 124.71
Nursing/Assisted Living NULL
BB 25.09
Career  71.03
Convenience Store    43.56
Cafeteria 8.78
Mobile Unit 11.37
Alcohol Manufacturer 0.02
Feed/Animal Products 4.04
Church   0.06
Vending  0.05
Maple Products 0.0
Food Manufacturer  NULL
Juice/Cider 0.13
Coffee Roaster 0.0
Distribution 0.0
Private Club 11.86
Dairy    356.83
Home-based Business 12.83
Meat/Poultry/Seafood 0.18
Hospital 13.43
Restaurant 1365.78
Senior meals 1.1
Food Establishment  NULL
Lodging 183.06
Camp    27.54
Sweets   2.41
Residential Facility 32.12
Concession Stand 7.38
0.37    NULL
Food Processor 13.5
Time taken: 0.791 seconds, Fetched 36 row(s)
spark-sql>

```

Streaming Dataset 1:

Query 1: SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1
 where year=2021
 group by food_type
 order by Sum_tons_Surplus limit 1;

```

sharathreddy1240@g15foodwaste-m:~$ spark-sql
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
ivysettings.xml file not found in HIVE_HOME or $HIVE_CONF_DIR, /etc/hive/conf.dist/ivysettings.xml will be used
23/11/30 05:39:40 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/11/30 05:39:40 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/11/30 05:39:40 INFO org.apache.spark.SparkEnv: Registering BlockManagerHeartbeat
23/11/30 05:39:40 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
Spark master: yarn, Application Id: application_1701316582184_0007
spark-sql> show tables
  >;
default foodwaste_streaming_1  false
default foodwaste_static_1  false
Time taken: 0.347 seconds, Fetched 2 row(s)
spark-sql> SELECT food_type, round (sum (tons_surplus_due_to_cause), 2) as Sum_tons_Surplus FROM foodwaste_streaming_1
   > where year=2021
   > group by food type
   > order by Sum_tons_Surplus limit 1;
23/11/30 05:39:23 WARN org.apache.hadoop.hive.ql.session.SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory.
23/11/30 05:39:24 WARN org.apache.hadoop.util.concurrent.ExecutorHelper: Thread (Thread[GetFileInfo #1,5,min]) interrupted:
java.lang.InterruptedException
   at com.google.common.util.concurrent.AbstractFuture.get(AbstractFuture.java:510)
   at com.google.common.util.concurrent.FluentFuture$TrustedFuture.get(FluentFuture.java:88)
   at org.apache.hadoop.util.concurrent.ExecutorHelper.logThrowableFromAfterExecute(ExecutorHelper.java:48)
   at org.apache.hadoop.util.concurrent.HadoopThreadPoolExecutor.afterExecute(HadoopThreadPoolExecutor.java:90)
   at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1157)
   at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
   at java.lang.Thread.run(Thread.java:750)
Breads & Bakery 2151409.52
Time taken: 11.345 seconds, Fetched 1 row(s)
spark-sql>

```

Query 2: SELECT sector, cause_group FROM foodwaste_streaming_1 group by sector, cause_group order by sector;

```

Welcome - Google Cloud console | BigQuery - ADTA5240F23G1 | ssh.cloud.google.com/v2/ssh | Downloads | Hive Analysis: Setup & Query ...
← → C ssh.cloud.google.com/v2/ssh/projects/adta5240f23g1/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024...
Downloads | Hive Analysis: Setup & Query ...
SSH-in-browser
at com.google.common.util.concurrent.FluentFuture$7$trustedFuture.get(FluentFuture.java:88)
at org.apache.hadoop.util.concurrent.ExecutorHelper.logThrowableFromAfterExecute(ExecutorHelper.java:48)
at org.apache.hadoop.util.concurrent.HadoopThreadPoolExecutor.afterExecute(HadoopThreadPoolExecutor.java:90)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1157)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
at java.lang.Thread.run(Thread.java:750)
Breads & Bakery 2151409.52
Time taken: 11.345 seconds, Fetched 1 row(s)
spark-sql> SELECT sector, cause_group FROM foodwaste_streaming_1
> group by sector, cause_group
> order by sector;
23/11/30 05:40:54 WARN org.apache.hadoop.util.concurrent.ExecutorHelper: Thread [Thread[GetFileInfo #1,5,main]] interrupted;
java.lang.InterruptedException
at com.google.common.util.concurrent.AbstractFuture.get(AbstractFuture.java:510)
at com.google.common.util.concurrent.FluentFuture$7$trustedFuture.get(FluentFuture.java:88)
at org.apache.hadoop.util.concurrent.ExecutorHelper.logThrowableFromAfterExecute(ExecutorHelper.java:48)
at org.apache.hadoop.util.concurrent.HadoopThreadPoolExecutor.afterExecute(HadoopThreadPoolExecutor.java:90)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1157)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
at java.lang.Thread.run(Thread.java:750)
Farm Buyer Rejections
Farm Not Harvested
Farm Spoiled
Farm Excess
Foodservice Spoiled
Foodservice Mistakes & Malfunctions
Foodservice Trimmings & Byproducts
Foodservice Date Label Concerns
Foodservice Excess
Foodservice Other
Manufacturing Trimmings & Byproducts
Manufacturing Excess
Manufacturing Buyer Rejections
Residential Trimmings & Byproducts
Residential Food Safety
Residential Other
Residential Date Label Concerns
Residential Mistakes & Malfunctions
Residential Excess
Residential Spoiled
Retail Food Safety
Retail Mistakes & Malfunctions
Retail Trimmings & Byproducts
Retail Spoiled
Retail Other
Retail Excess
Retail Date Label Concerns
sector cause_group
Time taken: 4.683 seconds, Fetched 28 row(s)
spark-sql> 

```

Query 3: `SELECT sector, round (sum(us_dollars_surplus_due_to_cause)) as Sum_usdollars_Surplus, round(sum(tons_surplus_due_to_cause))as Sum_tons_Surplus FROM foodwaste_streaming_1 group by sector;`

```

Welcome - Google Cloud console | BigQuery - ADTA5240F23G1 | ssh.cloud.google.com/v2/ssh | Downloads | Hive Analysis: Setup & Query ...
← → C ssh.cloud.google.com/v2/ssh/projects/adta5240f23g1/zones/us-central1-a/instances/g15foodwaste-m?authuser=1&hl=en_GB&projectNumber=1514872024...
Downloads | Hive Analysis: Setup & Query ...
SSH-in-browser
Linux g15foodwaste-m 5.10.0-0.debian10.16-cloud-amd64 #1 SMP Debian 5.10.127-2-bpo10+1 (2022-07-28) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Nov 30 05:24:54 2023 from 35.235.244.34
sh@rathreddy1240g15foodwaste-m:~$ spark-sql
Setting default log level to "WARN".
To alter the log level use sc.setLogLevel(newLevel).
IvySettings.xml file will be loaded from $HOME or HIVE_CONF_DIR, /etc/hive/conf.dist/ivysettings.xml will be used
23/11/30 05:42:07 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/11/30 05:42:08 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/11/30 05:42:08 INFO org.apache.spark.SparkEnv: Registering BlockManagerMasterHeartbeat
23/11/30 05:42:08 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
Spark master: yarn, Application Id: application_1701316582184_0008
spark-sql> SELECT sector, round (sum(us_dollars_surplus_due_to_cause)) as
> Sum_usdollars_Surplus, round(sum(tons_surplus_due_to_cause) ) as Sum_tons_Surplus
> FROM foodwaste_streaming_1
> group by sector;
23/11/30 05:42:54 WARN org.apache.hive.ql.session.SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory
Foodservice 1.620496708103E12 1.42792396E8
Farm 1.22411488956E11 1.99927161E8
Residential 2.266991075597E12 4.77928223E8
sector NULL NULL
Retail 2.67867236023E11 5.3634789E7
Manufacturing 3.97533834249E11 1.47095529E8
Time taken: 11.971 seconds, Fetched 6 row(s)
spark-sql> 

```

Query 4: `SELECT sub_sector, round (sum(us_dollars_surplus_due_to_cause)) as Sum_usdollars_Surplus, round(sum(tons_surplus_due_to_cause)) as Sum_tons_Surplus`

FROM foodwaste_streaming_1
group by sub_sector;

```
spark-sql> SELECT sub_sector, round (sum(us_dollars_surplus_due_to_cause)) as
  > Sum_usdolars_Surplus, round(sum(tons_surplus_due_to_cause)) as Sum_tons_Surplus
  > FROM foodwaste_streaming_1
  > group by sub_sector;SELECT sub_sector, round (sum(us_dollars_surplus_due_to_cause)) as
  > Sum_usdolars_Surplus, round(sum(tons_surplus_due_to_cause)) as Sum_tons_Surplus
  > FROM foodwaste_streaming_1
  > group by sub_sector;
sub_sector      NULL      NULL
Healthcare     6.2855770522E10 3002153.0
Full Service Restaurants 8.80834303794E11   6.612368E7
Military       7.339233106E9 1054331.0
College & University 2.535380048E10 3311955.0
Corrections    7.496702038E9 1068095.0
Recreation     7.2933734659E10 6812361.0
Refreshment Services 1.9750726E10 4153573.0
Other          4.996360422E9 688127.0
Bars And Taverns 3.719313407E9 375396.0
Caterers        3.4772003894E10 3100639.0
K-12 Education 5.3991150771E10 7512744.0
Business & Industry 2.5789910422E10 3573693.0
Non Profit      0.05480363482E12 8.78585702E8
Lodging         1.49542858548E11 1.199526787
Transportation 1.4368304923E10 1510925.0
Limited Service Restaurants 2.56732467113E11 2.8509455E7
Time taken: 1.098 seconds, Fetched 17 row(s)
sub_sector      NULL      NULL
Healthcare     6.2855770522E10 3002153.0
Full Service Restaurants 8.80834303794E11   6.612368E7
Military       7.339233106E9 1054331.0
College & University 2.535380048E10 3311955.0
Corrections    7.496702038E9 1068095.0
Recreation     7.2933734659E10 6812361.0
Refreshment Services 1.9750726E10 4153573.0
Other          4.996360422E9 688127.0
Bars And Taverns 3.719313407E9 375396.0
Caterers        3.4772003894E10 3100639.0
K-12 Education 5.3991150771E10 7512744.0
Business & Industry 2.5789910422E10 3573693.0
Non Profit      0.05480363482E12 8.78585702E8
Lodging         1.49542858548E11 1.199526787
Transportation 1.4368304923E10 1510925.0
Limited Service Restaurants 2.56732467113E11 2.8509455E7
Time taken: 0.804 seconds, Fetched 17 row(s)
spark-sql> 
```

Query 5: SELECT cause_group, year, round (sum (tons_surplus_due_to_cause),2)as Sum_tons_Surplus FROM foodwaste_streaming_1
group by year, cause_group order by year, Sum_tons_Surplus;

```
Excess 2016 2.002875256E7
Trimmings & Byproducts 2016 2.452260465E7
Mistakes & Malfunctions 2017 1381765.4
Buyer Rejections 2017 1583550.14
Other 2017 2336438.7
Food Safety 2017 2555961.85
Date Label Concerns 2017 6168906.0
Spoiled 2017 1.300186713E7
Not Harvested 2017 1.393783579E7
Excess 2018 2.039040462E7
Trimmings & Byproducts 2017 2.426260986E7
Mistakes & Malfunctions 2018 1488229.4
Buyer Rejections 2018 1581622.15
Other 2018 2460917.37
Food Safety 2018 2733045.79
Date Label Concerns 2018 6571231.81
Spoiled 2018 1.298139134E7
Not Harvested 2018 1.396419584E7
Excess 2019 2.041414141E7
Trimmings & Byproducts 2018 2.525651769E7
Mistakes & Malfunctions 2019 1559282.89
Buyer Rejections 2019 1594635.82
Other 2019 2429637.7
Food Safety 2019 2755419.52
Date Label Concerns 2019 6607050.61
Spoiled 2019 1.3214407E7
Not Harvested 2019 1.389348044E7
Excess 2019 2.368214892E7
Trimmings & Byproducts 2019 2.565506948E7
Buyer Rejections 2020 1590676.32
Mistakes & Malfunctions 2020 1651396.37
Other 2020 2652567.04
Food Safety 2020 3082402.12
Date Label Concerns 2020 6623992.22
Not Harvested 2020 1.33265871E7
Spoiled 2020 1.428692235E7
Excess 2020 1.942853985E7
Trimmings & Byproducts 2020 2.74542879E7
Mistakes & Malfunctions 2021 1450434.46
Buyer Rejections 2021 1616169.35
Other 2021 2557772.14
Food Safety 2021 2958992.54
Date Label Concerns 2021 7007652.92
Not Harvested 2021 1.272039109E7
Spoiled 2021 1.383422158E7
Excess 2021 2.156001158E7
Trimmings & Byproducts 2021 2.726553455E7
cause_group      year      NULL
Time taken: 1.983 seconds, Fetched 109 row(s)
spark-sql> 
```

Streaming Dataset 2:

Query 1: SELECT * FROM foodwaste_streaming_2
order by Food_Waste_Collected desc

limit 10;

```
spark-sql> SELECT * FROM foodwaste_streaming_2
  > order by Food_Waste_Collected desc
  > limit 10;
23/11/30 16:13:53 WARN org.apache.hadoop.hive.ql.session.SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory
Collection Date Food Waste Collected   Estimated Earned Compost Created
2023-07-17T00:00:00Z    998.4   100
2023-05-30T00:00:00Z    991.4   99
2023-06-26T00:00:00Z    988.4   99
2022-12-27T00:00:00Z    987.4   99
2023-04-12T00:00:00Z    985.4   99
2022-09-22T00:00:00Z    979.4   98
2022-09-21T00:00:00Z    978.8   98
2022-10-01T00:00:00Z    968.4   97
2022-10-31T00:00:00Z    953.4   95
Time taken: 10.026 seconds, Fetched 10 row(s)
spark-sql> show tables;
default foodwaste_streaming_1  false
default foodwaste_streaming_2  false
default foodwastestatic_1  false
Time taken: 0.323 seconds, Fetched 3 row(s)
spark-sql> ■
```

Query 2: `SELECT * FROM foodwaste_streaming_2 where Collection_Date like '2023-06-19%';`

Here is the **time comparison chart** of Hive and Spark while running queries on both the platforms.

Static Dataset:

Query	Timetaken_Hive (in seconds)	Timetaken_Spark (in seconds)
1	7.841	17.583
2	6.066	1.586
3	8.369	0.704
4	6.839	1.174
5	6.73	0.791

Streaming Dataset 1:

Query	Timetaken_Hive (in seconds)	Timetaken_Spark (in seconds)
1	6.905	11.345
2	6.842	4.683
3	6.776	11.971
4	17.103	0.804
5	6.521	1.983

Streaming Dataset 2:

Query	Timetaken_Hive (in seconds)	Timetaken_Spark (in seconds)
1	14.26	10.026
2	6.302	-

References:

- Berman, D. (n.d.). Comparing Apache Hive vs. Spark. *logz.io*. <https://logz.io/blog/hive-vs-spark/>
- Google Cloud. (n.d.). BigQuery overview. *BigQuery*.
<https://cloud.google.com/bigquery/docs/introduction>
- HBS Online. (2021, February 2). *8 Steps in the data life cycle*.
<https://online.hbs.edu/blog/post/data-life-cycle>
- Leal Filho, W., Ribeiro, P.C.C., Setti, A.F.F. et al. (2023, May 6). Toward food waste reduction at universities. *Environment, Development and Sustainability*.
<https://doi.org/10.1007/s10668-023-03300-2>
- Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F. (2017, May 24). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing* (239). 29-57.
<https://doi.org/10.1016/j.neucom.2017.01.078>