

Analyzing consumer shopping behavior from a large multi-category online store



Introduction & Background

- The advent of technology and the internet has led to an increasing demand for E-commerce industries.
- People of all ages prefer online shopping and increase in trend of online shopping was highly noticeable during the pandemic and post pandemic
- The tremendous growth in online shopping has created a need for the business to understand how the customer perceives online shopping and purchases to improve their overall business performance.
- Large-scale data analytics in e-commerce is still in its nascent stage and there is plenty to learn about it in all facets.
- Analyzing such data which records the online shopper user activity is known as consumer shopping behavior analysis.



Goals & Objectives

- To do an in-depth analysis of the ecommerce dataset to gain insights on customer behavior and activity which can help business make strategic decisions.
- To understand big data analytics and how the evolving technologies help deal with such huge volume of data and leverage it to extract meaningful insights.



Project application & Impact

- The consumer shopping behavior analysis is the process of discovering, interpreting, and communicating data patterns and insights related to E-commerce online business.
- It helps in measuring the user behavior, market and performance trends, and ROI. This analysis can be used by the business to show ROI for the campaigns and make better decisions to reduce costs, increase sales, and make business improvements accordingly.

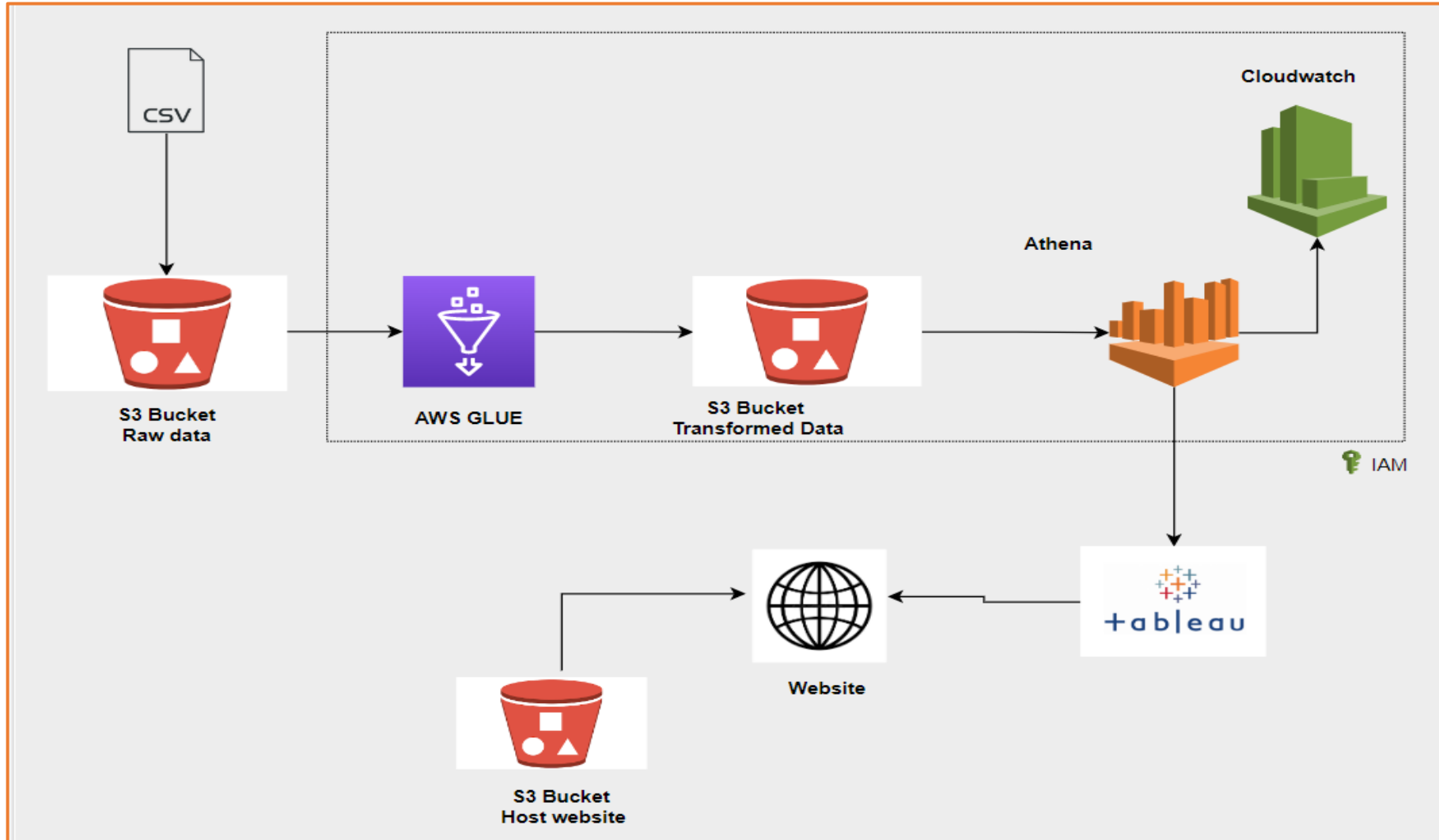


Data Description

- Data Source: Kaggle
- Dataset from a multcategory e-commerce online store
- Data Size: 15GB
- Data period: October-November 2019
- Dataset comprises of attributes such as product ID, product category, product sub-category, price, user details, user session details, brand, event_type, etc describing the user shopping activity.
- Link: <https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store>

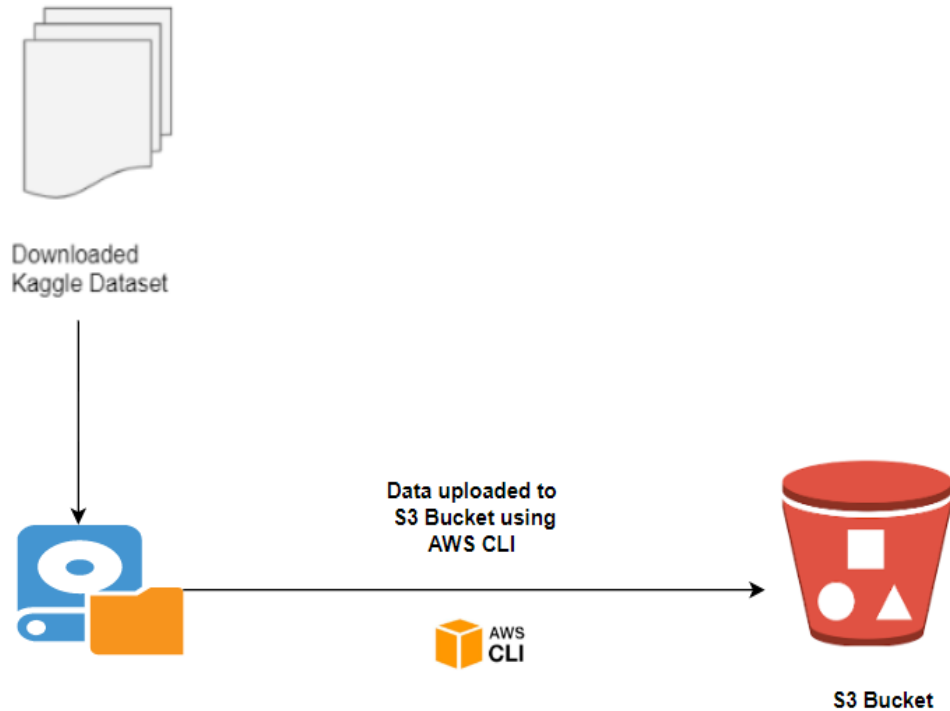


System Design/ ELT Architecture

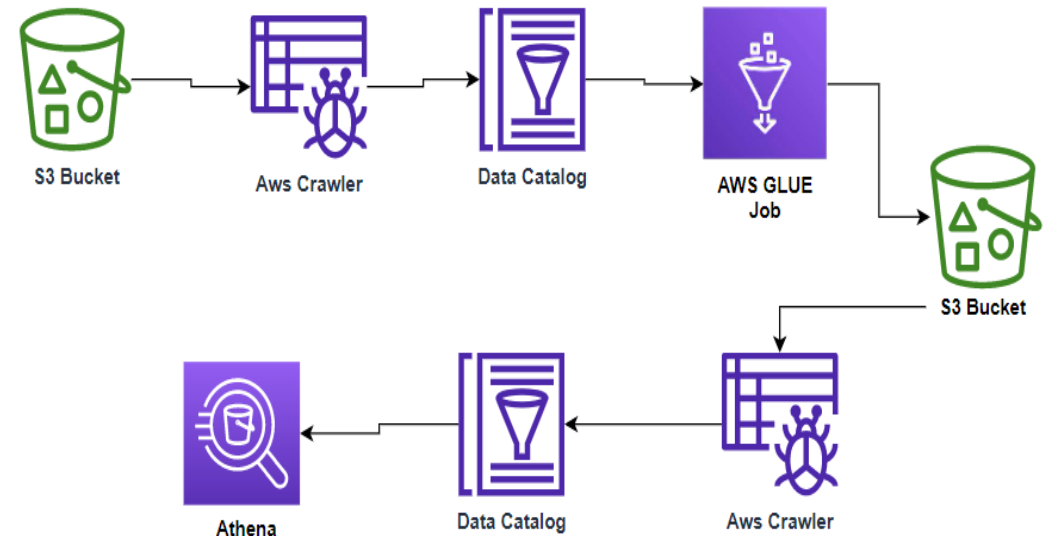


Data Loading And Transformation

Loading Data Into S3

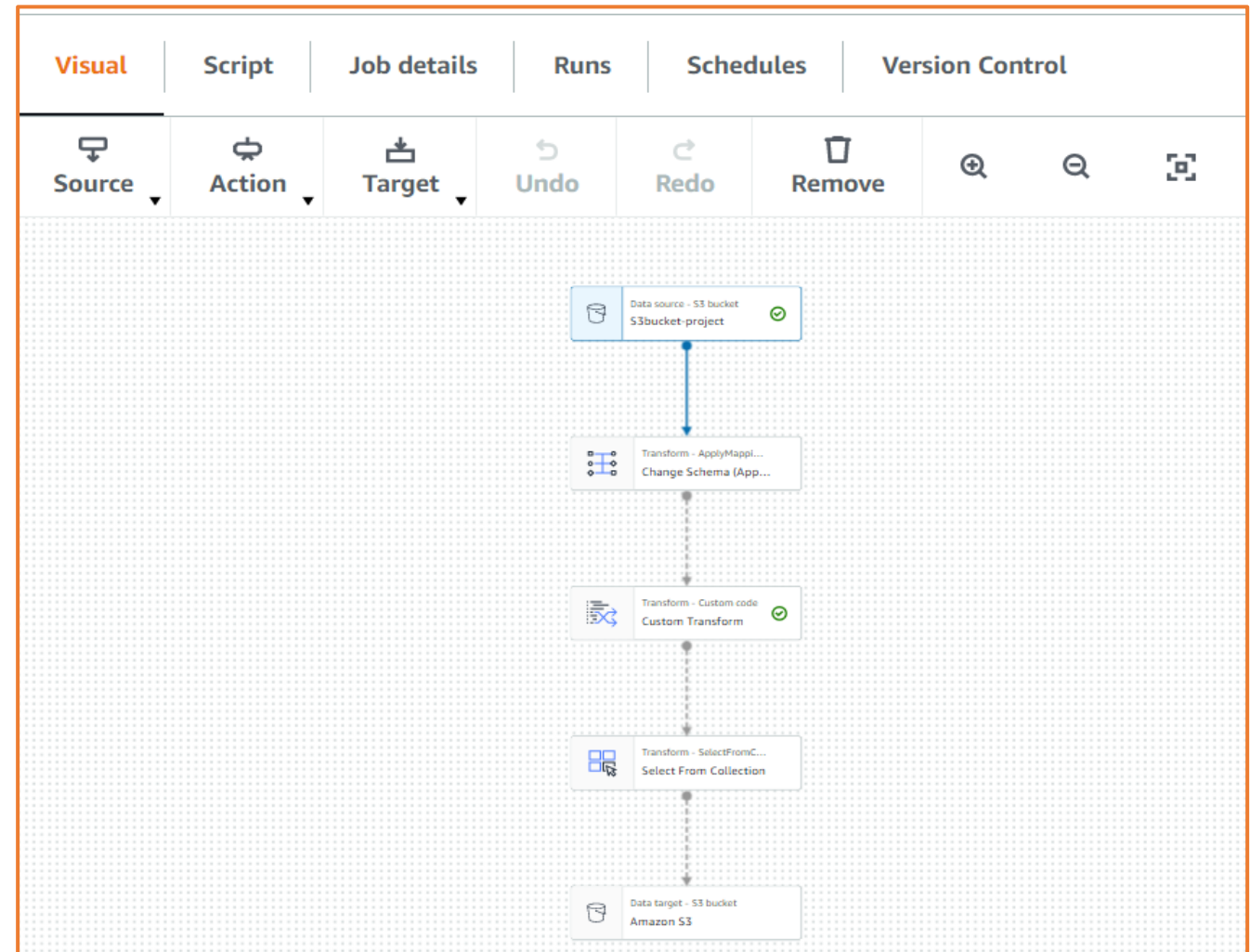


Data Transformation



ETL – Glue Job

- The columns are modified with relevant datatypes.
- The column category_code is split into three columns.



Data Analysis

- Athena is serverless.
- In Athena the data in S3 bucket is analyzed using SQL.

Querying Data in Athena

The screenshot displays the Amazon Athena Query Editor interface. At the top, the breadcrumb navigation shows 'Amazon Athena > Query editor'. Below this, there are tabs for 'Editor', 'Recent queries', 'Saved queries', and 'Settings'. The 'Workgroup' is set to 'primary'. The query editor shows a list of queries, with 'Query 5' selected and marked as completed. The SQL query is: `Select event_time, category_1, category_2, category_3 from projectoutput.etlprojectoutput where event_type = 'view'`. Below the query, there are buttons for 'Run again', 'Explain', 'Cancel', 'Clear', and 'Create'. A 'Reuse query results' toggle is also present. The 'Query results' tab is active, showing a green 'Completed' status bar with details: 'Time in queue: 171 ms', 'Run time: 3 min 26.052 sec', and 'Data scanned: 18.02 GB'. Below this, there are 'Copy' and 'Download results' buttons. A search bar for rows is provided. The results are displayed in a table with columns: '#', 'event_time', 'category_1', 'category_2', and 'category_3'. Two rows of data are visible.

#	event_time	category_1	category_2	category_3
1	"2019-11-03 01:28:51.0"	electronics	video	tv
2	"2019-11-03 01:28:51.0"	computers	notebook	

Connection To Tableau

- Tableau desktop is used to create connections with Athena for creating dashboards.

Connecting Tableau to Athena

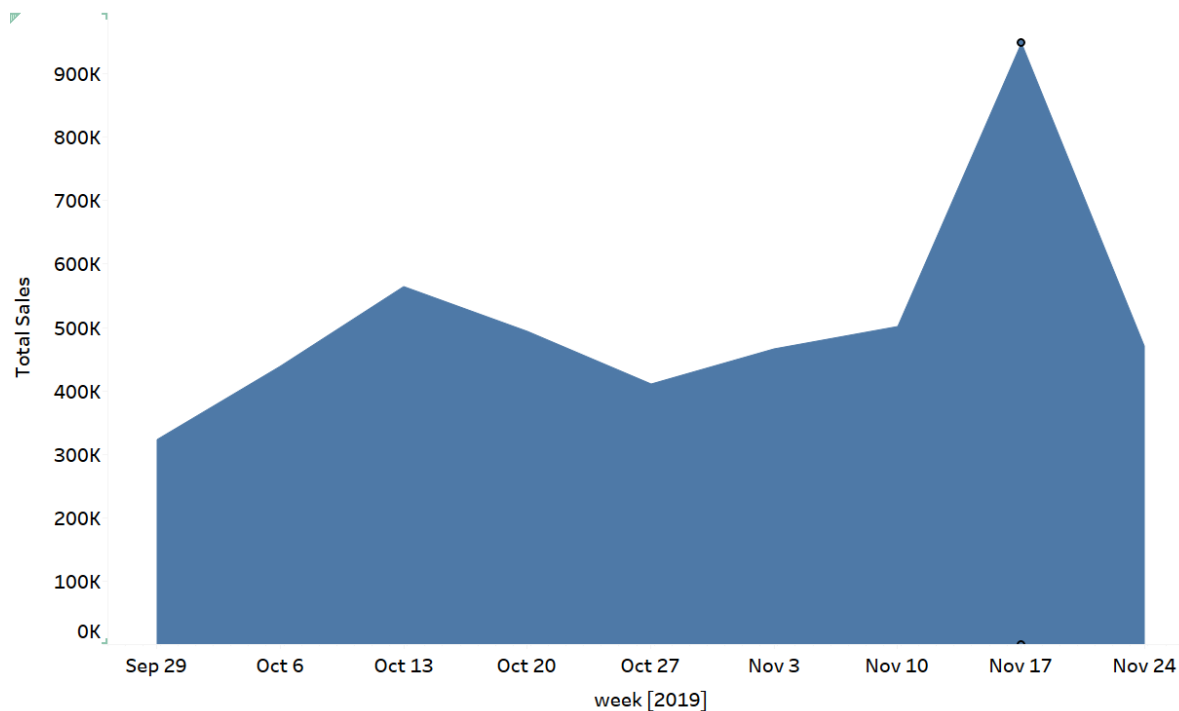
The screenshot shows the 'Amazon Athena' connection window in Tableau. The 'General' tab is selected, displaying the following configuration details:

- Server:** athena.us-east-1.amazonaws.com
- Port:** 443
- S3 Staging Directory:** s3://bigdataprojectbucket/projectoutput/
- Access Key ID:** AKIA2VBEFWX6WKHN6YXO
- Secret Access Key:** [Redacted with dots]

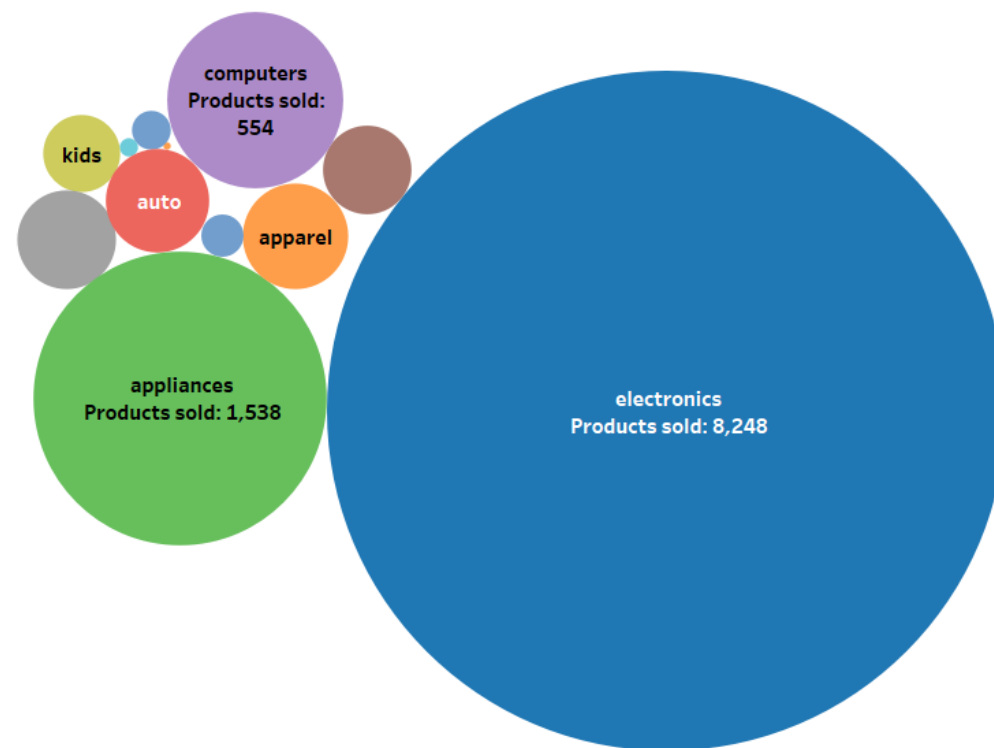
A 'Sign In' button is located at the bottom right of the window.

Data Visualizations

Total Sales per week

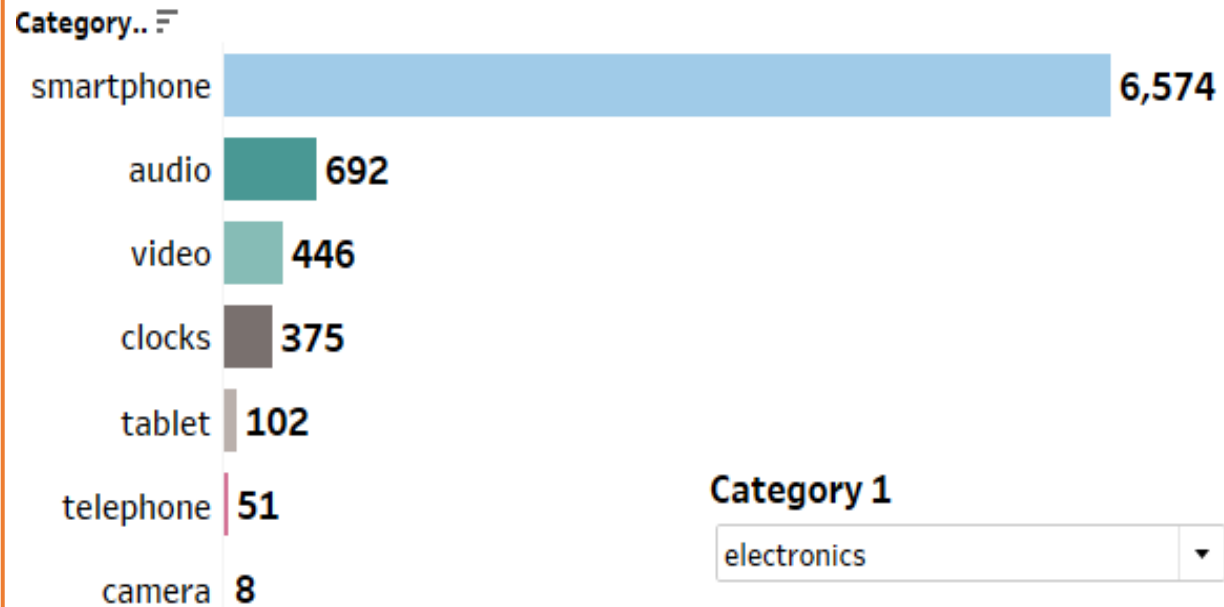


Top Sold Product Category

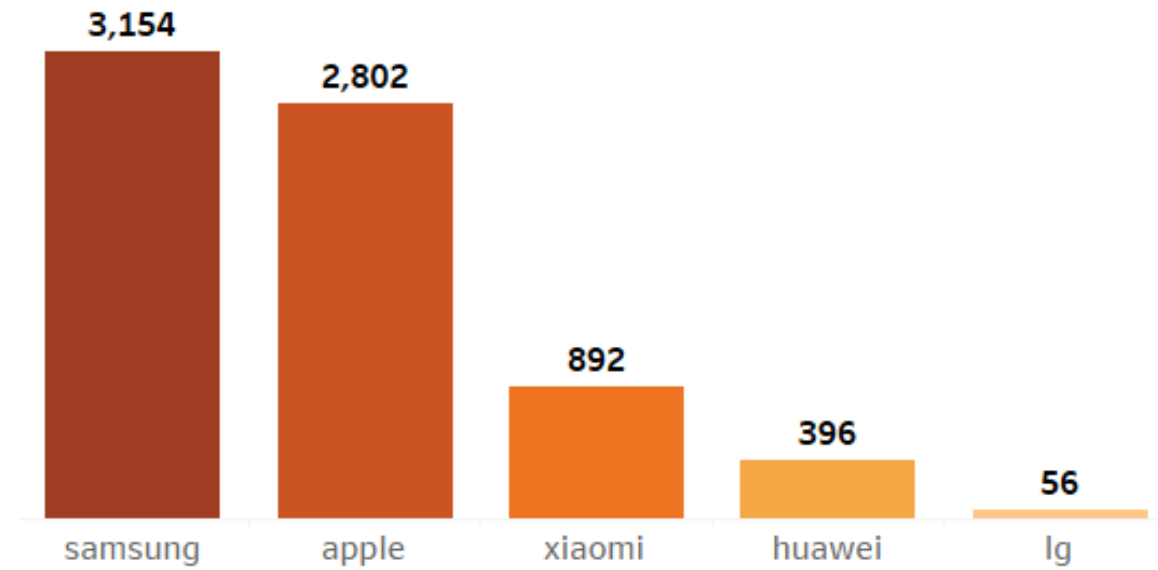


Data Visualizations

Top Sold Product Sub-category for Electronics

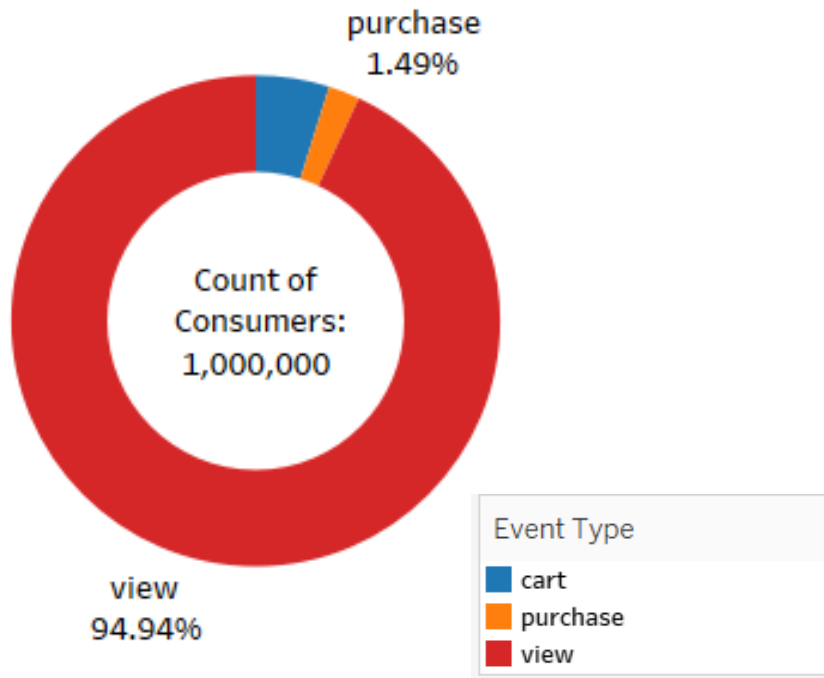


Popular Brand

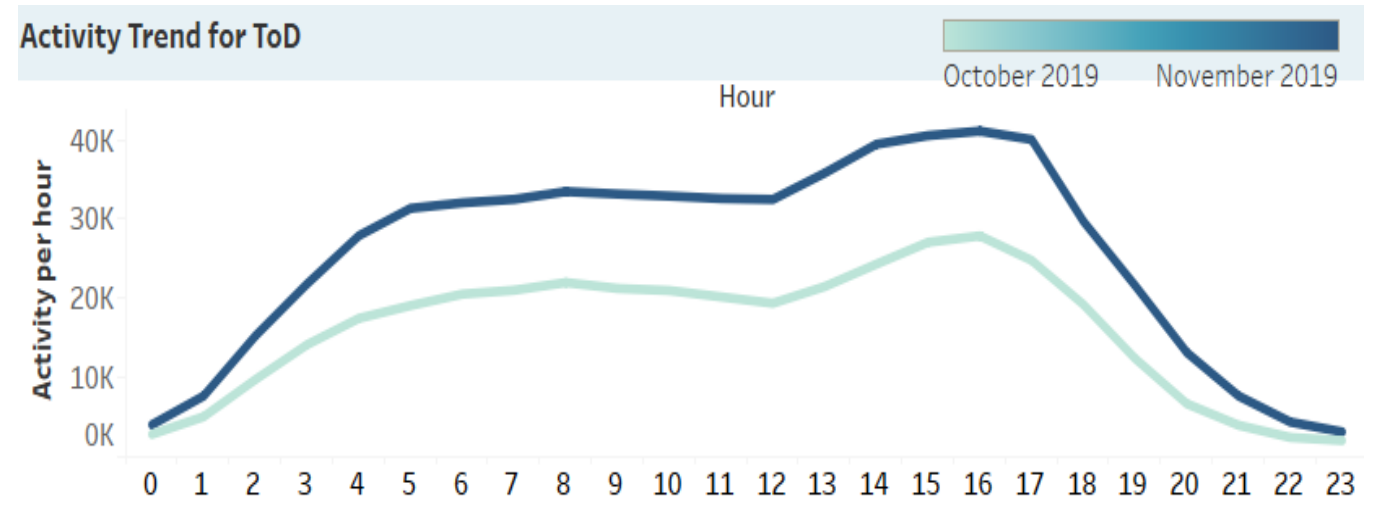


Data Visualizations

Distribution for each event



Activity Trend throughout the day

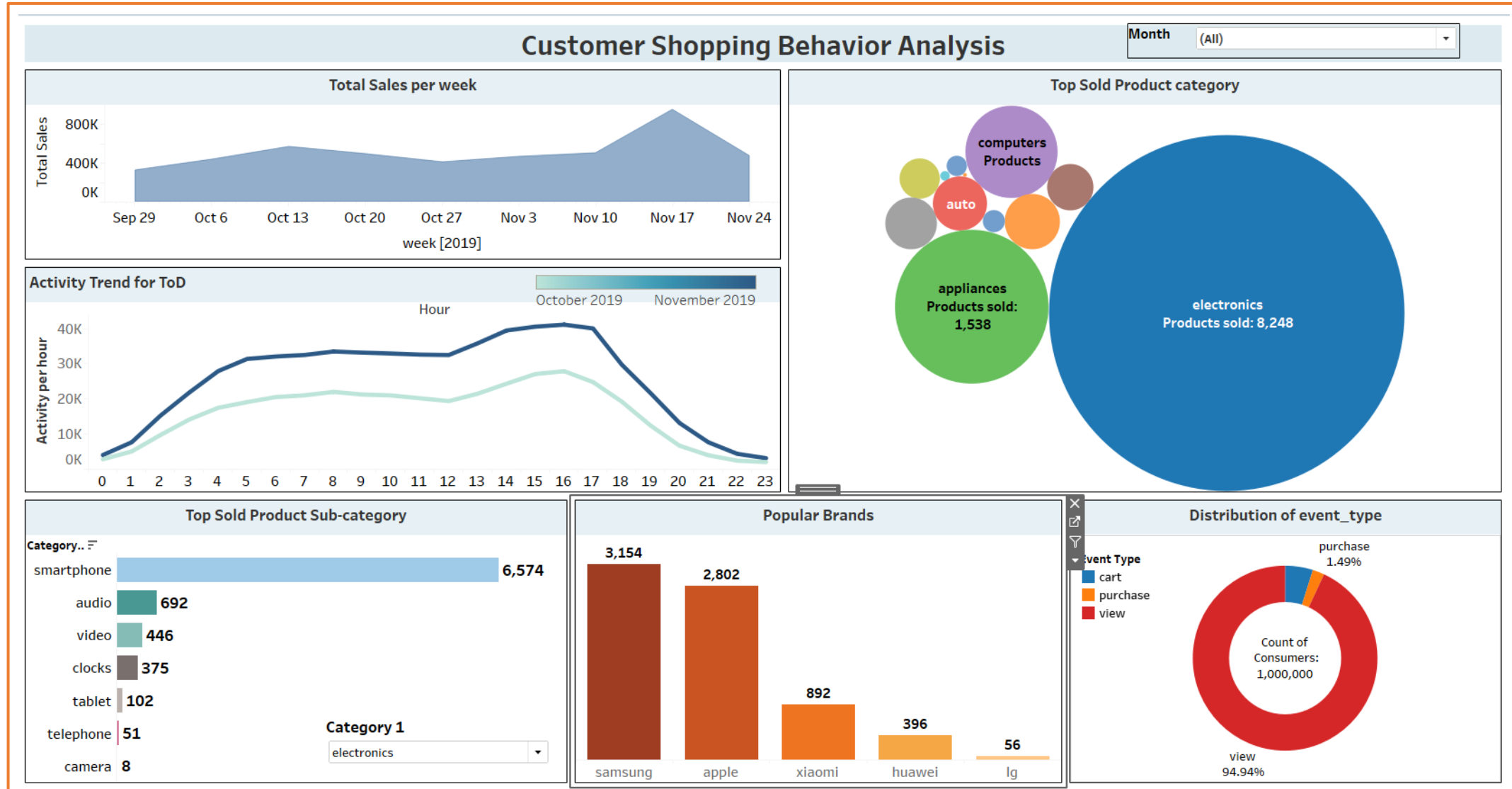


Data Visualizations

Month-Over-Month Growth %

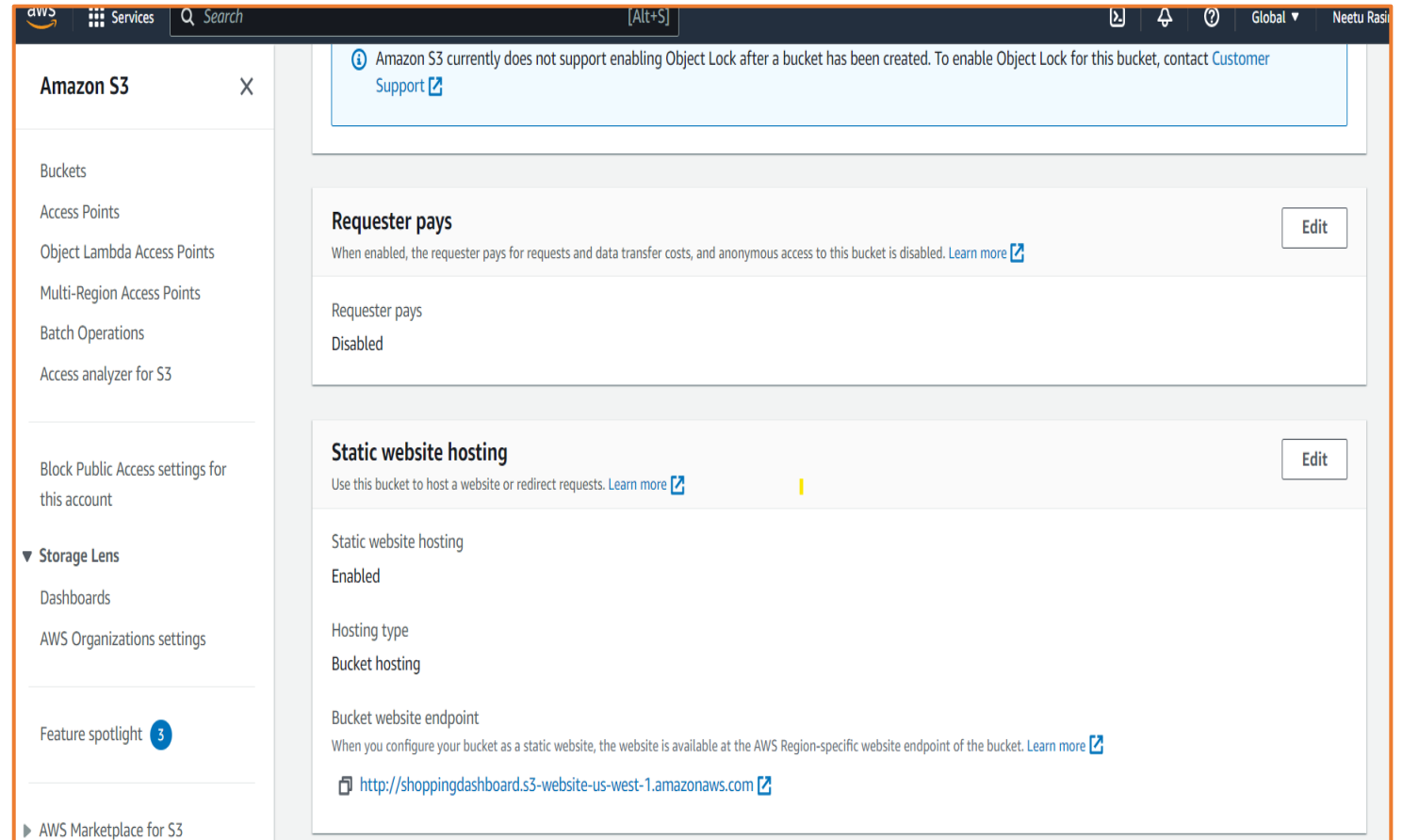
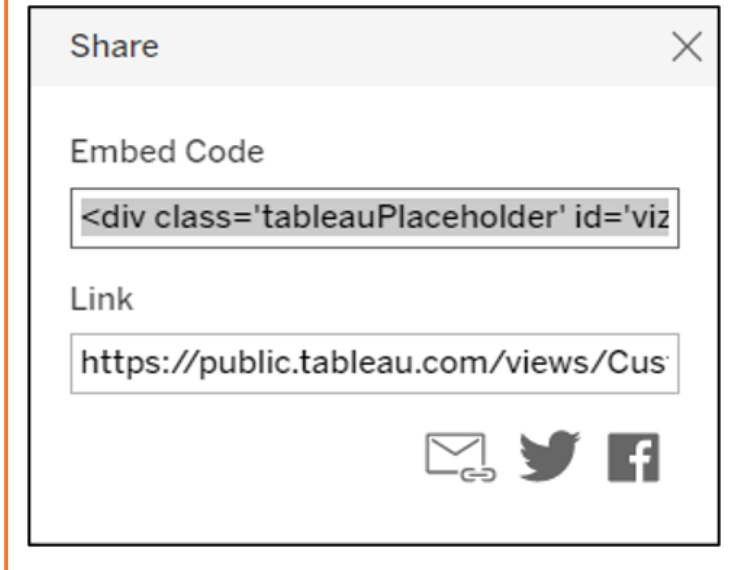
Category 1	Count of Products		Event_Time Earnings		Percentage Change	
	October	November	October	November	October	November
apparel	71	128	\$6K	\$10K	0.00%	72.82%
appliances	678	860	\$119K	\$167K	0.00%	40.60%
auto	95	96	\$13K	\$13K	0.00%	-3.25%
computers	243	311	\$97K	\$134K	0.00%	38.39%
construction	72	68	\$10K	\$7K	0.00%	-32.33%
electronics	3,839	4,409	\$1,643K	\$1,856K	0.00%	12.92%
furniture	74	100	\$16K	\$20K	0.00%	25.26%
kids	53	53	\$9K	\$8K	0.00%	-7.67%

Dashboard



Web Integration

Embedded Code In Tableau



Static website link: <https://shoppingdashboard.s3.us-west-1.amazonaws.com/index.html>

Conclusion

- Analyzed the most active hour and day for online shopping
- Observed the most purchased product category and sub-category by most users
- Observed the most popular brands for each product category.
- Consumer experience can be improved by ensuring that there is no server crash during the most active period.
- Offering discounts on products during weekdays can attract more users for shopping through the week further contributing to increase in sales.



Future Work

- Analyze Consumer shopping behavior for a couple of years. (High Volume data)
- Use machine learning algorithms for predicting the demands for products based on the consumer shopping behavior for inventory planning.
- Build recommendation systems to recommend the products to the users based on their past purchases and predicting their future purchases.



Image References

- <https://www.analyticsvidhya.com/blog/2015/06/start-journey-kaggle/>
- <https://www.softcrylic.com/unleash-the-power-of-data-analytics-in-aws/>
- <https://jgmediallc.com/2018/11/24/e-commerce-is-the-solution-for-rowan-county-nc/>
- <http://revenueconsultant.com/2016/12/12/8-incredible-tips-to-increase-sales-performance/>
- <https://sigma.software/about/media/top-5-major-challenges-big-data-analytics-and-ways-tackle-them>
- <https://www.analyticsvidhya.com/blog/2015/06/start-journey-kaggle/>
- https://www.google.com/search?q=data+loading+from+aws+cli+unto+S3+bucket&tbm=isch&ved=2ahUKEwiZxee3bj7AhUSgl4IHTR0DiMQ2cCegQIABAA&oq=data+loading+from+aws+cli+unto+S3+bucket&gs_lcp=CgNpbWcQAzoECCMQJzoFCAAQgAQ6BAGAEb46BggAEAgQHjoHCAAQgAQQGfDrFFi6gAFglYcBaABwAHgAgAFSiAGYFJIBAJM3mAEAoAEbqgELZ3dzLXdpei1pbWfAAQE&sclient=img&ei=8h14Y5mEMJKAuvQPtOi5mAl&bih=714&biw=1536&rlz=1C1CHBF_enUS912US912&hl=en#imgsrc=J36yJMLwnqnIAM
- https://www.google.com/search?q=aws+cli+images&rlz=1C1CHBF_enUS912US912&sxsrf=ALiCzsbMXliCPLTN_ePDeor14XpdR6eidw:1668816454519&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj493f-bj7AhUeMDQIHd5DAsQ_AUoAXoECAEQAw&biw=1536&bih=714&dpr=1.25#imgsrc=7u79qN_DnoFTBM
- https://www.google.com/search?q=aws+glue+etl+job+icon&tbm=isch&ved=2ahUKEwjm9qb_h7n7AhVbiI4IHRPDC0kQ2cCegQIABAA&oq=aws+glue+etl+icon&gs_lcp=CgNpbWcQARgBMgUIABCABDlGCAAQbXAEogQIABBDogYIABAIEB5QuApYnBdgrChoAHAAeACAawCIAcAFkgECMTCYAQCgAQGqAQtd3Mtd2l6LWltZ8ABAQ&sclient=img&ei=Ni14YaDMtuQuvQPk4avyAQ&bih=714&biw=1519&rlz=1C1CHBF_enUS912US912&hl=en#imgsrc=FiZ-Tg2N9vWSEM
- draw.io
- <https://www.shutterstock.com/search/conclusions-icons>
- <https://www.shutterstock.com/search/future-work-icons>



Thank you