

# NIT2202 Group Assignment – Live Demonstration Report

---

## *~Introduction*

In this project, I built an end-to-end analytics pipeline to load, model, and visualize Greater Brisbane employment data using **Amazon S3**, **Snowflake**, and **Tableau Cloud**. I created a secure S3 bucket for the source CSV, defined Snowflake database objects (database, schema, tables, file format, stage), and ingested the data with COPY INTO. After validating the load with SQL (preview queries and aggregations), I connected Tableau Cloud **live** to Snowflake and produced three core views: a **Top 10 Industries** bar chart, a **Sub-sector Treemap**, and an **Industry Share** pie/donut. These visuals feed a single dashboard that lets me rank industries, drill into sub-sectors (e.g., Hospitals, Food & Beverage, Construction Services), and compare overall market share—highlighting Health Care & Social Assistance as the leading employer in 2021 and a tight cluster of mid-tier industries just below it. This approach demonstrates the full 18-point workflow required for the live technical demo, from cloud storage through to business insight.

## 1. Snowflake – Lab 04 Step 17 & 18 Results

The screenshot displays the Snowflake web interface. On the left, the database structure is shown: S8122353\_DB, INFORMATION\_SCHEMA, PUBLIC, S8122353\_SCHEMA, Tables, Stages, SNOWFLAKE, SNOWFLAKE\_LEARNING\_DB, SNOWFLAKE\_SAMPLE\_DATA, TB\_101, and USER\$S8122353. The main panel shows the SQL query:

```
29 FILE_FORMAT = (TYPE = 'CSV' SKIP_HEADER = 1);
30
31 COPY INTO industry_employment
32 FROM @s8122353_industry_employment_stage
33 FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = '\t' SKIP_HEADER = 1)
34 ON_ERROR = 'CONTINUE';
35
36 | SELECT * FROM industry_employment LIMIT 10;
37
38 COPY INTO industry_employment
39 FROM @s8122353_industry_employment_stage
40 FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = '\t' SKIP_HEADER = 1)
41 ON_ERROR = 'CONTINUE';
42
```

The results table shows 10 rows of data:

	SERIES	SA2_MAIN11	GCCSA_NAME	LOA_NAME_2013	SA4_NAME	SA3_NAME	SA2_NAME
1	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
2	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
3	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
4	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
5	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
6	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
7	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
8	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
9	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
10	POW	304041104	Greater Brisbane	Brisbane (C)	Brisbane - West	The Gap - Enoggera	Upper Kedi

Query Details: Query duration 175ms, Rows 10, Query ID 01be6e95-3204-bada-0-...

This screenshot shows a Snowflake web interface where SQL queries are being executed. The workspace displays a database named **S8122353\_DB** with its schema and tables. A **COPY INTO** command is used to load data from a stage into the **industry\_employment** table, with CSV file formats specified and error handling set to **CONTINUE**. Below, a **SELECT \* FROM industry\_employment LIMIT 10;** query retrieves the first ten rows of data.

The screenshot displays the Snowflake web interface. On the left, the database structure is shown: S8122353\_DB, INFORMATION\_SCHEMA, PUBLIC, S8122353\_SCHEMA, Tables, Stages, SNOWFLAKE, SNOWFLAKE\_LEARNING\_DB, SNOWFLAKE\_SAMPLE\_DATA, TB\_101, and USER\$S8122353. The main panel shows the SQL query:

```
48 WHERE YEAR = 2030
49 GROUP BY YEAR;
50
51
52
53 SELECT
54 SUM(SC2_2011) AS total_2011,
55 SUM(SC2_2013) AS total_2013,
56 SUM(SC2_2016) AS total_2016,
57 SUM(SC2_2021) AS total_2021,
58 SUM(SC2_2026) AS total_2026,
59 SUM(SC2_2031) AS total_2031,
60 SUM(SC2_2036) AS total_2036,
61 SUM(SC2_2041) AS total_2041
62 FROM industry_employment;
63
64
```

The results table shows 1 row of data:

	# TOTAL_2011	# TOTAL_2013	# TOTAL_2016	# TOTAL_2021	# TOTAL_2026	# TOTAL_2031	# TOTAL_2036
1	575570	573008	598816	622289	662500	712008	

Query Details: Query duration 401ms, Rows 1, Query ID 01be7187-3204-bbba-0-...

The query shown in the image performs aggregation using the **SUM()** function on multiple year-based columns from the table `industry_employment`. It calculates the total employment values for different years, including **2011, 2013, 2016, 2021, 2026, 2031, 2036, and 2041**.

The result set displays one row containing the total employment figures for each year. For example, in 2011 the total is **575,570**, in 2016 it rises to **598,816**, and by 2031 it reaches **712,008**. This query is useful for analyzing trends and growth in employment across future projections.

The screenshot shows the Snowflake web interface. On the left, the 'Databases' sidebar lists the schema structure, including 'S8122353\_DB' and 'INDUSTRY\_EMPLOYMENT'. The main panel displays a SQL query with line numbers 55 to 71. The query calculates the sum of employment for various years (2016, 2021, 2026, 2031, 2036, 2041) from the 'industry\_employment' table. Below the query, the 'Results' tab shows a table with two columns: 'GCCSA\_NAME' and '# EMPLOYMENT\_2021'. The first row shows 'Greater Brisbane' with a value of 622,289. To the right of the results, 'Query Details' indicate a duration of 75ms, 1 row, and a query ID. An 'Ask Copilot' button is also visible.

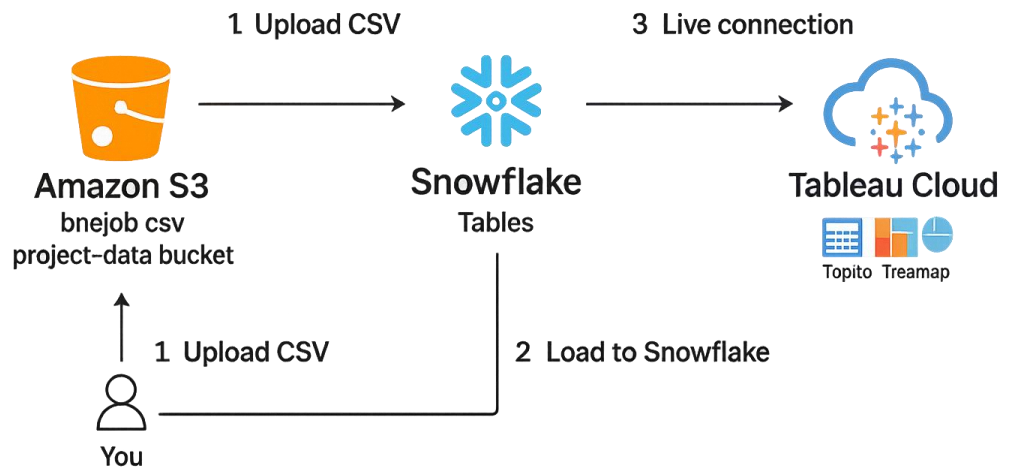
GCCSA_NAME	# EMPLOYMENT_2021
Greater Brisbane	622289

The query in the image calculates the total employment for the year **2021** grouped by region. Here, **GCCSA\_Name** represents the Greater Capital City Statistical Area, and the query sums employment data from the column `SC2_2021`. The results are grouped by each region and sorted in descending order of employment.

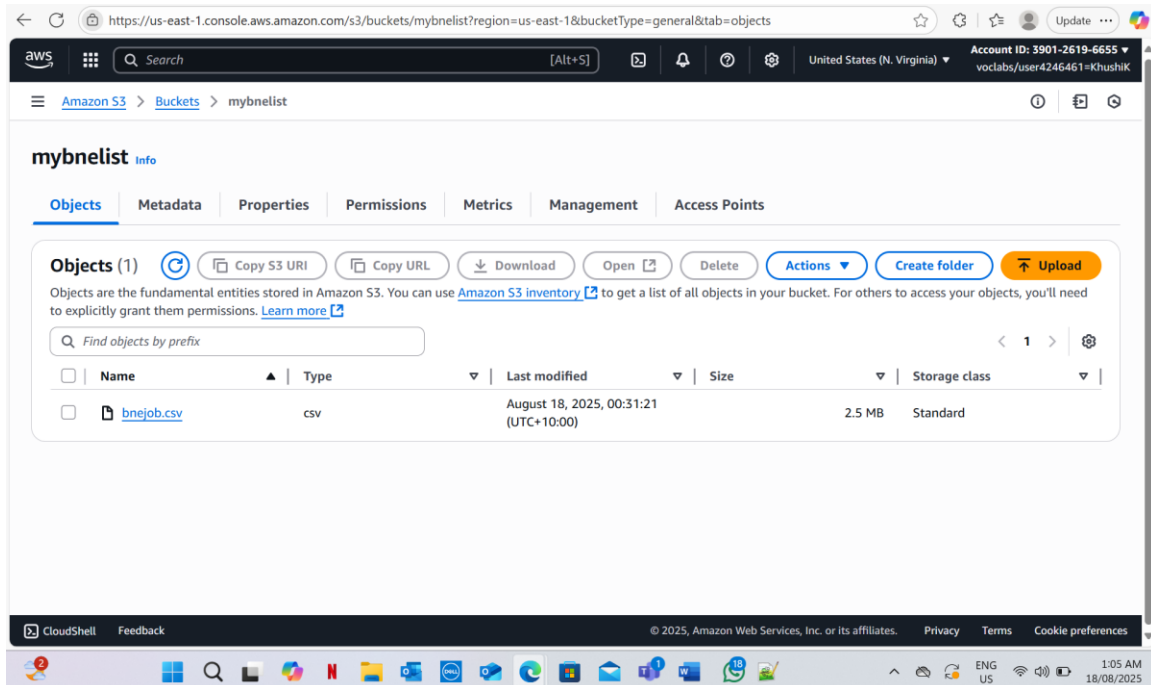
In the output shown, **Greater Brisbane** is listed with a total employment of **622,289** for 2021, making it the highest region in this dataset.

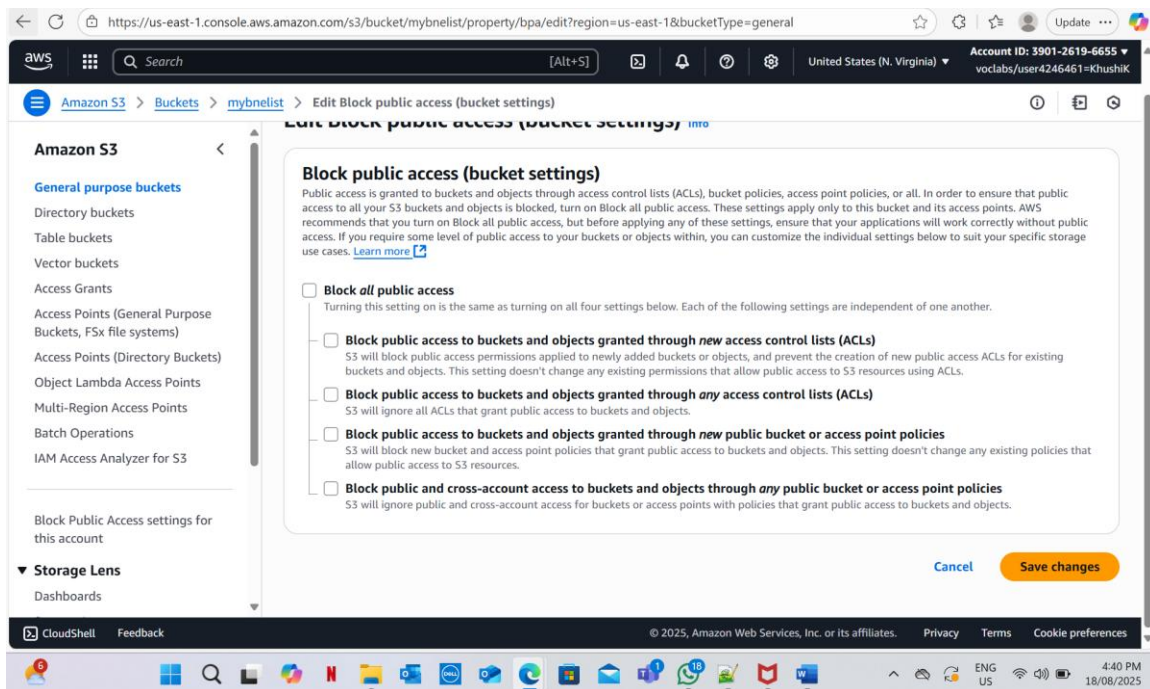
## 2. AWS Architecture – System Diagram

### NIT2022 – Data Flow: S3 → Snowflake – Tableau Cloud (2021)



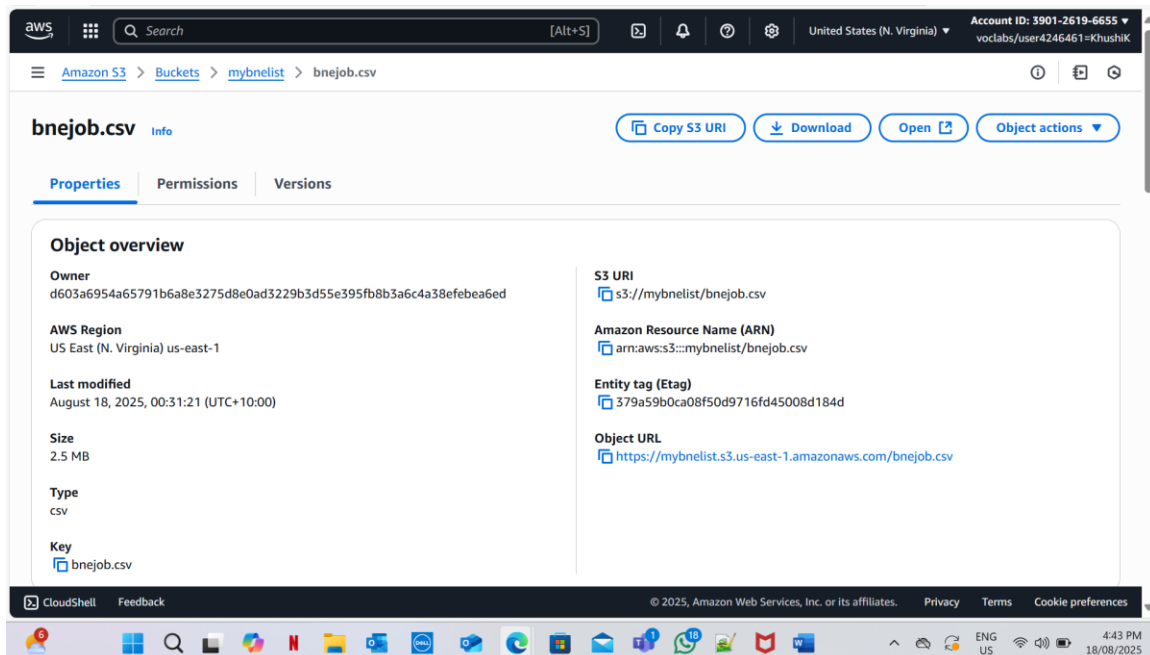
## 3. Amazon S3 – Bucket Creation





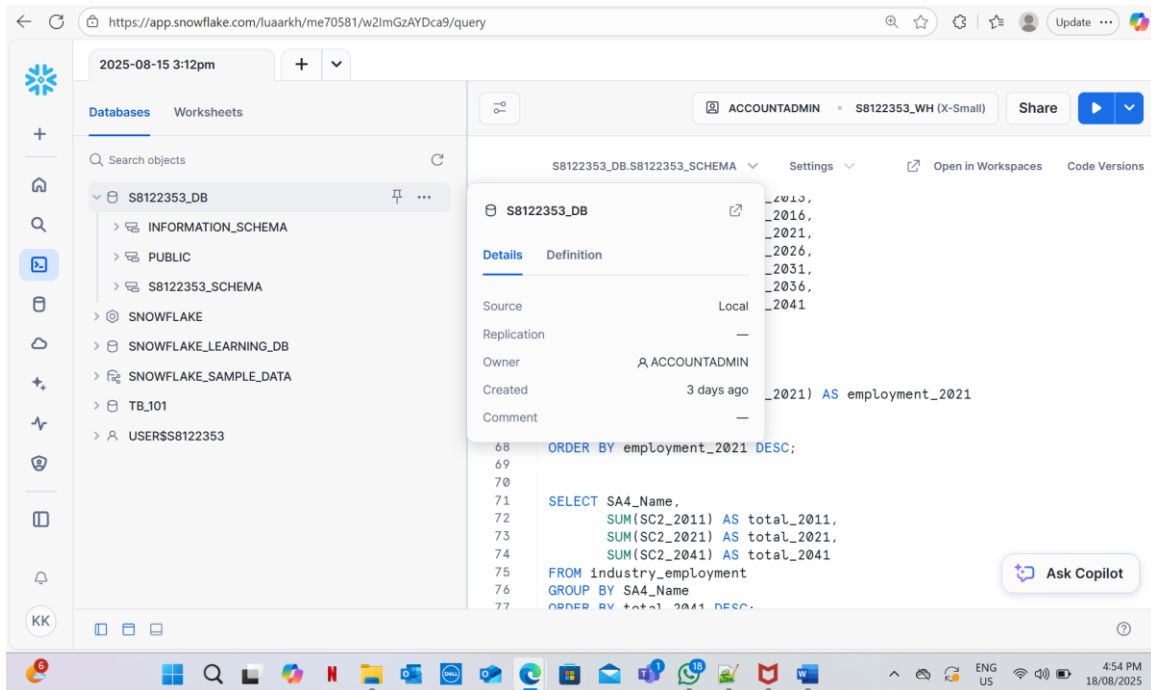
Amazon S3 bucket creation allows you to create a storage space to manage and organize files in the cloud. In the screenshots, a bucket named “mybnelist” was created. Inside this bucket, a file named **bnejob.csv** of size **2.5 MB** was uploaded under the Standard storage class. While creating and configuring the bucket, the **Block Public Access** settings were available to ensure security by restricting unauthorized access through ACLs and policies. This demonstrates how I created a secure S3 bucket, uploaded your dataset, and managed access control for safe storage.

#### 4. Data Source – Upload and Connect



The screenshot shows the **object details** of the uploaded data source file in Amazon S3. The file **bnejob.csv** was uploaded into the bucket **mybnelist**. It is a **CSV file** with a size of **2.5 MB**, last modified on **August 18, 2025**. This has a unique **S3 URI**, **Amazon Resource Name (ARN)**, and **Object URL**, which can be used to connect this data source with other AWS services such as Snowflake, Athena, or Redshift. These details ensure that the uploaded dataset is properly stored and accessible for further integration and analysis.

## 5. Snowflake – Database Setup



**CREATE DATABASE S8122353\_DB;**

## 6. Snowflake – Schema Definition

The screenshot shows the Snowflake web interface. On the left, the 'Databases' tab is active, and the 'S8122353\_SCHEMA' is selected under the 'S8122353\_DB'. The main panel displays the 'Definition' tab for the schema, showing a list of tables and their definitions. The tables listed are:

- SC2\_2011
- SC2\_2013
- SC2\_2016
- SC2\_2021
- SC2\_2026
- SC2\_2031
- SC2\_2036
- SC2\_2041

The 'Details' tab for the schema is also visible, showing the parent database as 'S8122353\_DB', the owner as 'ACCOUNTADMIN', and the creation time as '3 days ago'.

Create schema `s8122353_schema`;

## 7. Snowflake – Table Creation

The screenshot shows the Snowflake web interface. On the left, the 'Databases' tab is active, and the 'S8122353\_SCHEMA' is selected under the 'S8122353\_DB'. The main panel displays the 'Definition' tab for the schema, showing a list of tables and their definitions. The tables listed are:

- SC2\_2011
- SC2\_2013
- SC2\_2016
- SC2\_2021
- SC2\_2026
- SC2\_2031
- SC2\_2036
- SC2\_2041

The 'Details' tab for the schema is also visible, showing the parent database as 'S8122353\_DB', the owner as 'ACCOUNTADMIN', and the creation time as '3 days ago'.

In this step, I created a table named **industry\_employment** inside my schema **S8122353\_SCHEMA** in Snowflake. The table was defined with multiple columns to store employment-related data. Some columns, such as **Series**, **GCCSA\_Name**, **LGA\_NAME\_2013**,



*SA4\_Name, SA3\_Name, SA2\_Name, and Name fields, were created as VARCHAR to hold text information. Other columns, such as SC2\_2011, SC2\_2013, SC2\_2016, SC2\_2021, SC2\_2026, SC2\_2031, SC2\_2036, and SC2\_2041, were created as NUMBER to store employment counts for different years. After creating the table, I ran a **SELECT \* FROM industry\_employment;** query to confirm that the empty table was successfully created.*

## 8. Snowflake – Load Project Data

The screenshot displays the Snowflake web interface. The left sidebar shows the database structure with the 'industry\_employment' table under the 'S8122353\_SCHEMA' database. The main pane shows the SQL query executed: `CREATE OR REPLACE STAGE 's8122353_industry_employment_stage' URL = 's3://mybnelist/bnejob.csv' FILE_FORMAT = (TYPE = 'CSV', SKIP_HEADER = 1);` followed by `COPY INTO industry_employment`. The 'Results' pane shows a single row with the status '100% filled'. The 'Query Details' pane shows the query duration as 1.0s and the query ID as '01be7285-3204-bada-0...'. The '25 Queries' pane shows a list of recent queries, with the first query being the 'CREATE OR REPLACE STAGE' query.

*In this step, I created a **stage** in Snowflake to connect my table with the CSV file stored in Amazon S3.*

*This stage acts as a reference point to access the file **bnejob.csv** directly from the S3 bucket. I defined the file format as **CSV** and skipped the header row to avoid loading column names as data. The query executed successfully, as shown in the results, confirming that the stage was created and ready for data loading.*



The screenshot displays the Snowflake web interface with a query executed. The query is as follows:

```
COPY INTO industry_employment
FROM @S8122353_industry_employment_stage
FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = ',' SKIP_HEADER = 1)
ON_ERROR = 'CONTINUE';
SELECT * FROM industry_employment LIMIT 100;
```

The results table shows the following data:

	SERIES	SA2_MAIN11	GCCSA_NAME	LGA_NAME_1
1	POW	305041137	Greater Brisbane	Brisbane (C)
2	POW	305041137	Greater Brisbane	Brisbane (C)
3	POW	305041137	Greater Brisbane	Brisbane (C)
4	POW	305041137	Greater Brisbane	Brisbane (C)
5	POW	305041137	Greater Brisbane	Brisbane (C)
6	POW	305041137	Greater Brisbane	Brisbane (C)
7	POW	305041137	Greater Brisbane	Brisbane (C)
8	POW	305041137	Greater Brisbane	Brisbane (C)
9	POW	305041137	Greater Brisbane	Brisbane (C)
10	POW	304041104	Greater Brisbane	Brisbane (C)
11	POW	304041104	Greater Brisbane	Brisbane (C)
12	POW	304041104	Greater Brisbane	Brisbane (C)
13	POW	304041104	Greater Brisbane	Brisbane (C)

The interface also shows a list of 25 queries on the right, with the current query highlighted. The query details on the right indicate a query duration of 118ms and 100 rows returned.

*In this step, I loaded the data from my **S3 stage** into the **industry\_employment** table using the **COPY INTO** command. The code I executed was:*

*This command copied the contents of the CSV file **bnejob.csv** from my S3 bucket into the Snowflake table. I also used **ON\_ERROR = 'CONTINUE'** to ensure the load process would continue even if some rows had issues.*

*The result confirmed that the first 100 rows of data were successfully inserted, showing values like **Series (POW)**, **SA2\_MAIN11**, and **Greater Brisbane** region details. This verified that my data pipeline from S3 to Snowflake worked correctly.*

## 9. Snowflake – Preview Data

The screenshot shows the Snowflake web interface. The top navigation bar includes the Snowflake logo, a search bar, and user information (ACCOUNTADMIN, S8122353\_WH). The left sidebar shows a tree view of the database structure, including S8122353\_DB, INFORMATION\_SCHEMA, PUBLIC, S8122353\_SCHEMA, Tables, Stages, SNOWFLAKE, SNOWFLAKE\_LEARNING\_DB, SNOWFLAKE\_SAMPLE\_DATA, TB\_301, and USER\$S8122353. The main panel displays a SQL query in the editor:

```
29 FILE_FORMAT = (TYPE = 'CSV' SKIP_HEADER = 1);
30
31 COPY INTO industry_employment
32 FROM @S8122353_industry_employment_stage
33 FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = '\t' SKIP_HEADER = 1)
34 ON_ERROR = 'CONTINUE';
35
36 | SELECT * FROM industry_employment LIMIT 10;
37
38 COPY INTO industry_employment
39 FROM @S8122353_industry_employment_stage
40 FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = '\t' SKIP_HEADER = 1)
41 ON_ERROR = 'CONTINUE';
42
```

Below the query editor, the 'Results' tab is active, showing a table with 10 rows and 7 columns. The columns are labeled: SERIES, SA2\_MAIN11, GCCSA\_NAME, LGA\_NAME\_2013, SA4\_NAME, SA3\_NAME, and SA2\_NAME. The rows contain data for various locations, including Greater Brisbane, Brisbane (C), Brisbane Inner City, Brisbane Inner - West, and Toowong.

	SERIES	SA2_MAIN11	GCCSA_NAME	LGA_NAME_2013	SA4_NAME	SA3_NAME	SA2_NAME
1	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
2	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
3	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
4	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
5	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
6	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
7	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
8	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
9	POW	305041137	Greater Brisbane	Brisbane (C)	Brisbane Inner City	Brisbane Inner - West	Toowong
10	POW	304041104	Greater Brisbane	Brisbane (C)	Brisbane - West	The Gap - Enoggera	Upper Kedj

On the right side of the results table, there is a 'Query Details' panel showing the query duration (175ms), the number of rows (10), and the query ID (01be6e95-3204-bada-0-). There is also an 'Ask Copilot' button.

In this step, I ran the query to check if the data was successfully loaded into my **industry\_employment** table. The command I used was:

```
SELECT * FROM industry_employment LIMIT 10;
```

This query retrieves the first 10 rows from the table and displays all columns. The results confirmed that the dataset had been inserted properly, showing fields such as **Series (POW)**, **SA2\_MAIN11**, **GCCSA\_Name (Greater Brisbane)**, **LGA\_Name\_2013 (Brisbane C)**, **SA4\_Name (Brisbane Inner City)**, **SA3\_Name (Brisbane Inner - West)**, and **SA2\_Name (Toowong)**.

## 10. Snowflake SQL – Simple Query

The screenshot shows the Snowflake SQL interface. The left sidebar displays the database structure, including the 'INDUSTRY\_EMPLOYMENT' table. The main query editor contains the following SQL code:

```
33 COPY INTO industry_employment
34 FROM @S8122353_industry_employment_stage
35 FILE_FORMAT = (TYPE = 'CSV' FIELD_DELIMITER = ',' SKIP_HEADER = 1)
36 ON_ERROR = 'CONTINUE';
37 SELECT * FROM industry_employment LIMIT 100;
```

The 'Results' tab shows a table with 13 rows and 5 columns: SERIES, SA2\_MAIN11, GCCSA\_NAME, LGA\_NAME, and a fourth unnamed column. The 'Query Details' panel on the right indicates a query duration of 118ms and 100 rows returned. The '25 Queries' list on the far right shows the current query as the most recent.

## 11. Snowflake SQL – Advanced Query 1

The screenshot shows the Snowflake SQL interface with an advanced query. The SQL code in the editor is:

```
64 SELECT GCCSA_Name, SUM(SC2_2021) AS employment_2021
65 FROM industry_employment
66 GROUP BY GCCSA_Name
67 ORDER BY employment_2021 DESC;
68
69
70
71 SELECT SA4_Name,
72 SUM(SC2_2011) AS total_2011,
73 SUM(SC2_2021) AS total_2021,
74 SUM(SC2_2041) AS total_2041
75 FROM industry_employment
76 GROUP BY SA4_Name
77 ORDER BY total_2041 DESC;
78
79
80
```

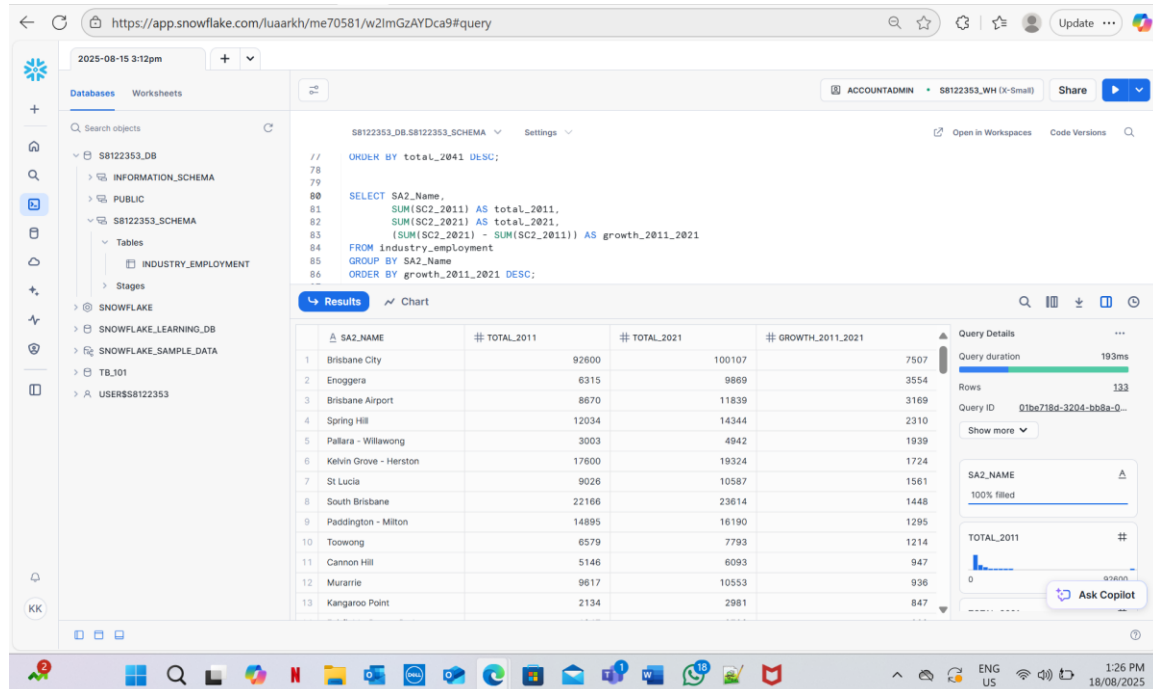
The 'Results' tab displays a table with 6 rows and 4 columns: SA4\_NAME, # TOTAL\_2011, # TOTAL\_2021, and # TOTAL\_2041. The 'Query Details' panel shows a query duration of 525ms and 6 rows returned. The '25 Queries' list on the right shows the current query as the most recent.

In this step, I performed an **aggregation query** to analyze employment data across different regions (SA4 areas) and years.

This query groups the dataset by **SA4\_Name** (regional areas) and calculates the total employment for the years **2011, 2021, and 2041**.

The output shows regions like **Brisbane Inner City, Brisbane South, Brisbane North, Brisbane West, Brisbane East, and Ipswich** with their respective totals. For example, **Brisbane Inner City** had 244,814 jobs in 2011, 265,563 in 2021, and is projected to reach 358,580 in 2041. This helped me compare historical data with future employment projections by region.

## 12. Snowflake SQL – Advanced Query 2

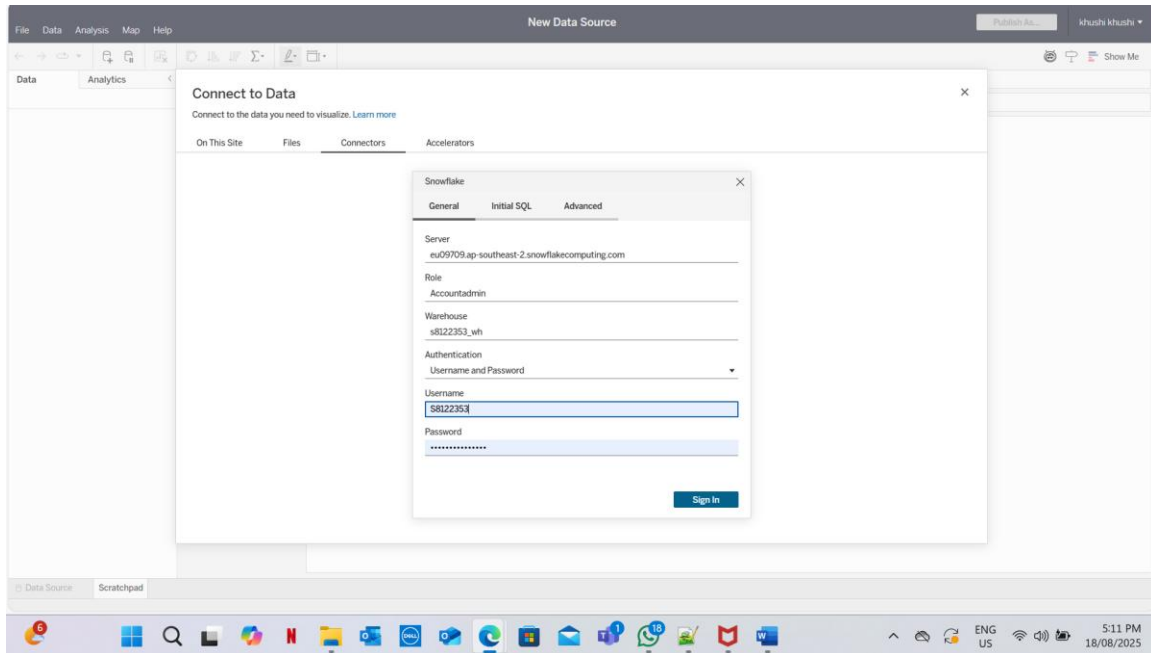


In this step, I wrote a query to calculate the **employment growth between 2011 and 2021** at the **SA2\_Name** level.

This query sums up the employment data for 2011 and 2021, then calculates the difference to show growth over that decade.

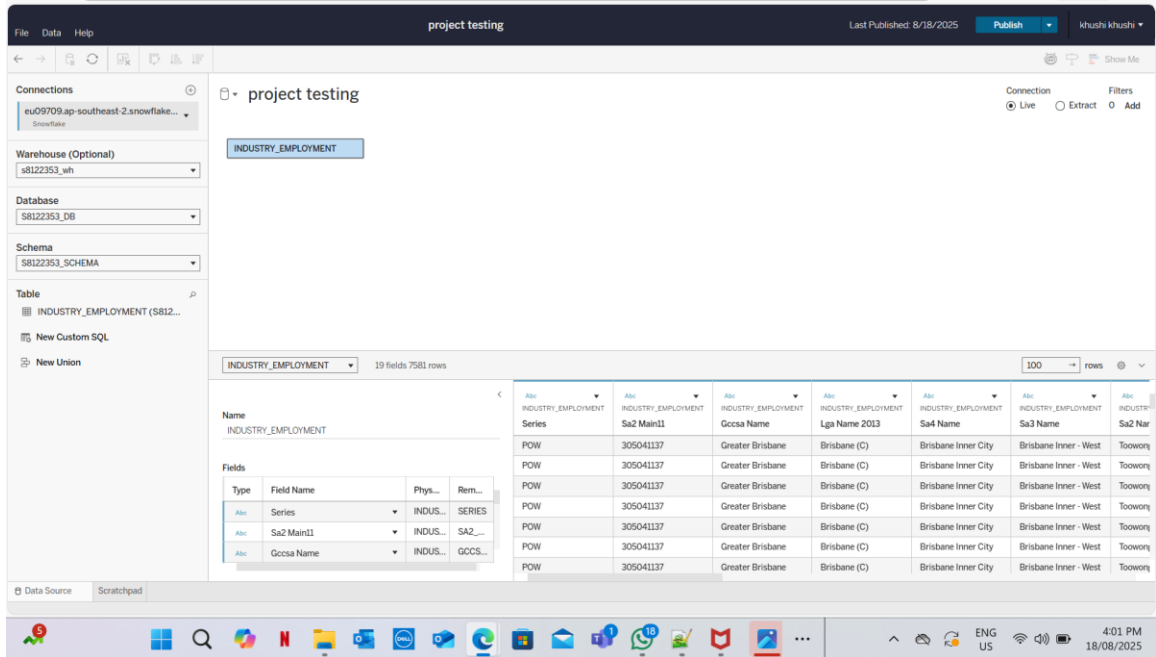
The output shows regions like **Brisbane City, Enoggera, Brisbane Airport, and Spring Hill**, along with their employment totals and growth. For example, **Brisbane City** had 92,600 jobs in 2011, which increased to 100,107 in 2021, giving a growth of **7,507 jobs**. This analysis highlights which local areas experienced the highest employment growth.

### 13. Tableau Cloud – Live Data Connection

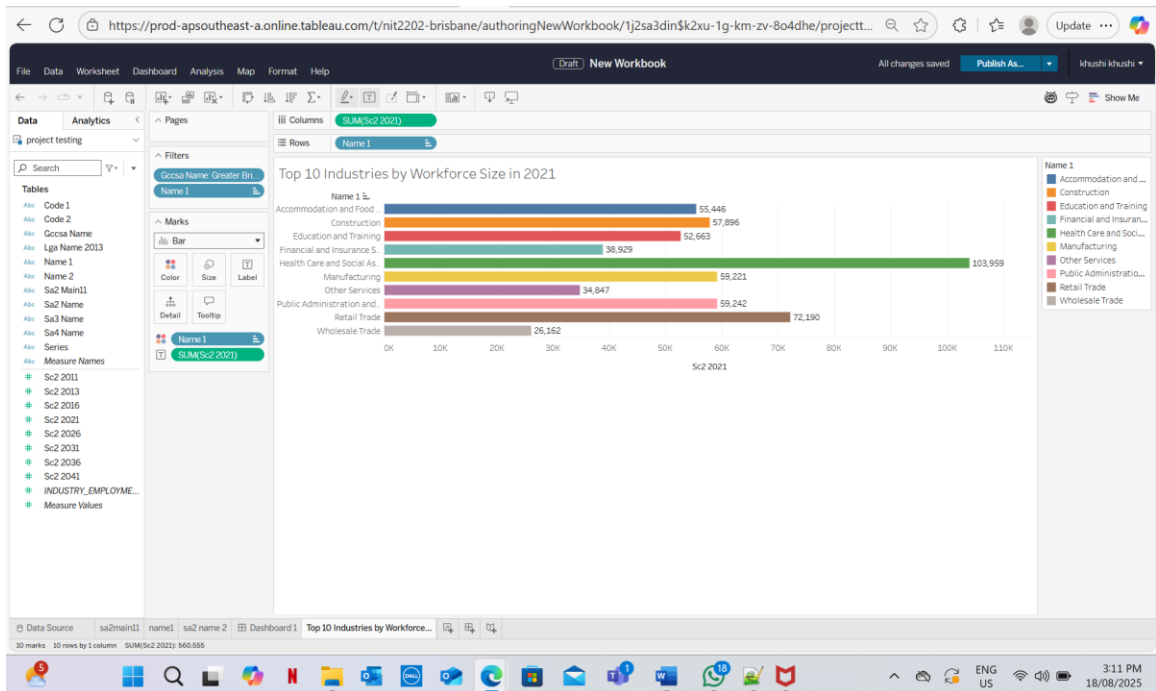


*In this screenshot, I connected **Tableau** to my **Snowflake database** to visualize the data. While setting up the connection, I selected the **Snowflake connector** and entered the details.*

*After filling in these details, I clicked **Sign In**. This successfully established a connection between Tableau and my Snowflake database, allowing me to import the **industry\_employment** table and create visualizations from the loaded employment dataset.*



## 14. Data Visualization – Chart 1

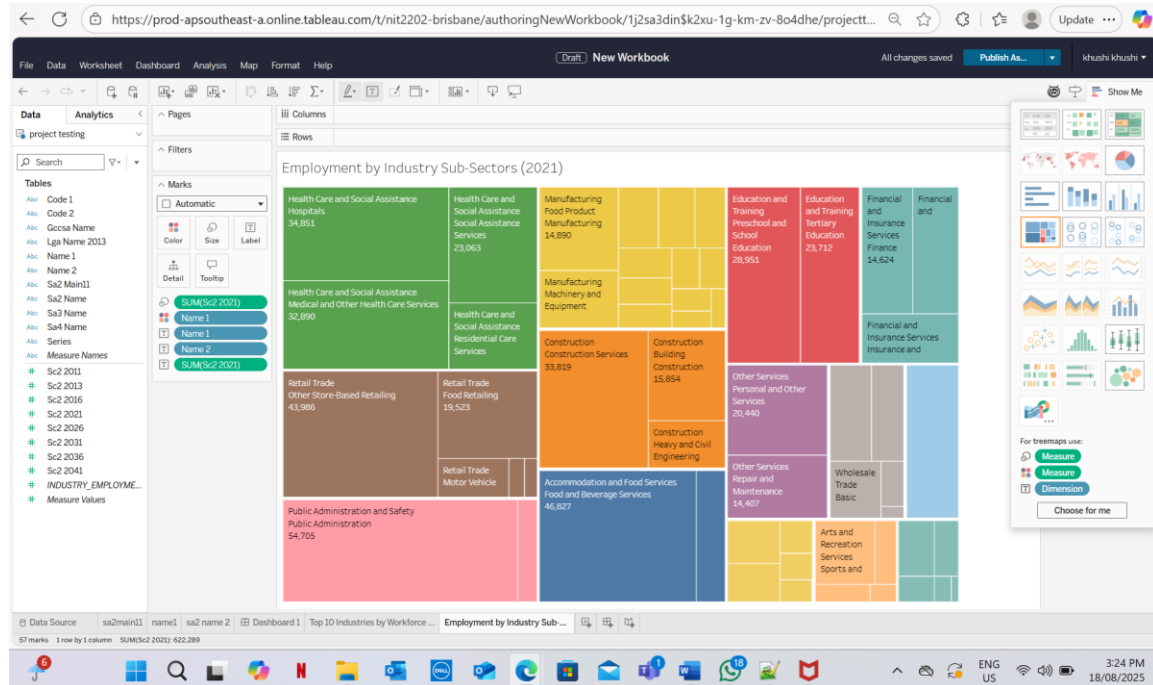


This chart shows the **top 10 industries by workforce size in 2021** for Greater Brisbane, using **SUM(Sc2 2021)**. I've placed industries (Name 1) on the Y-axis and total jobs on the X-axis, with colors to distinguish industries and labels showing counts. **Health Care & Social Assistance** is the largest at **103,959** jobs, followed by **Retail Trade (72,190)**. **Public Administration & Safety (59,242)** and **Manufacturing (59,221)** are almost tied, then **Construction (57,996)**, **Accommodation & Food Services (55,446)**, **Education & Training**



(52,663), **Financial & Insurance Services** (38,929), **Other Services** (34,847), and **Wholesale Trade** (26,162). Altogether, these ten industries account for **560,555** jobs. This view highlights healthcare as the clear leader and a middle cluster around 55–60k.

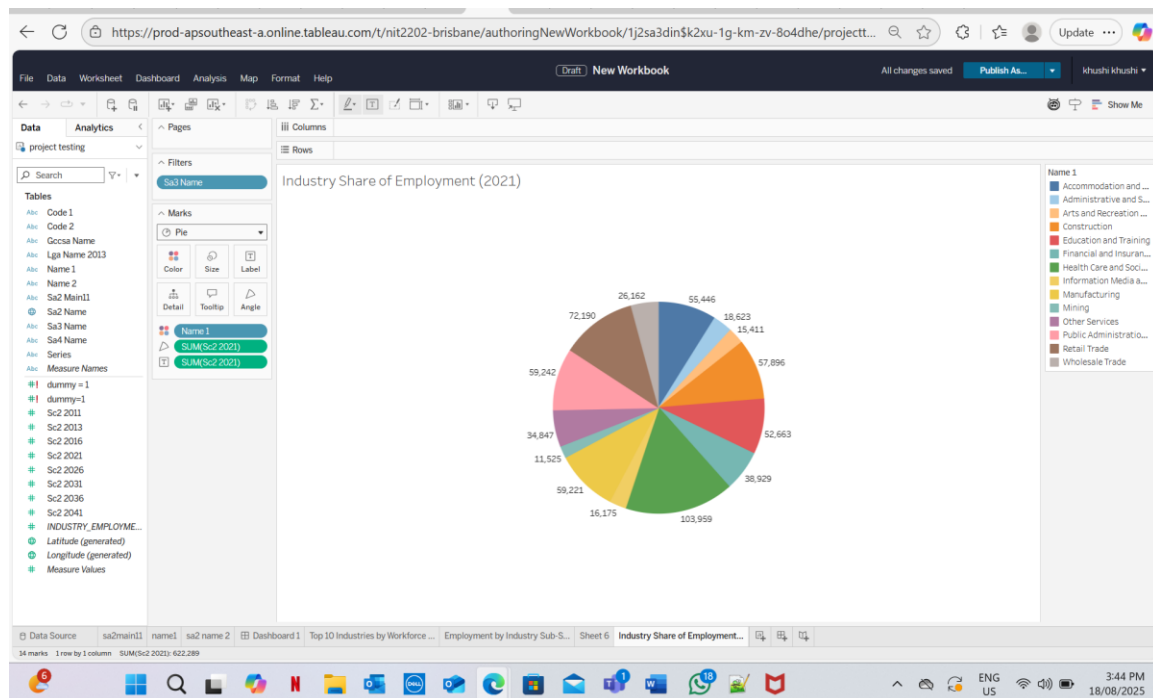
## 15. Data Visualization – Chart 2



This treemap shows **employment by industry sub-sectors in 2021**. Each rectangle is a sub-sector (Name 2), grouped and coloured by its parent **industry (Name 1)**, and the **size reflects total jobs (SUM(Sc2 2021))**. The biggest tiles highlight where jobs concentrate: **Public Administration** (~54.7k) stands out, along with **Food & Beverage Services** in **Accommodation & Food Services** (~46.8k) and **Other Store-Based Retailing** in **Retail** (~44.0k). Within **Health Care & Social Assistance**, employment is spread across large sub-sectors such as **Hospitals** (~34.9k), **Medical & Other Health Care Services** (~32.9k), **Residential Care Services**, and **Social Assistance Services** (~23.1k). **Construction Services** (~33.8k) is another major block, while **Education** splits across **Preschool & School Education** (~29.0k) and **Tertiary Education** (~23.7k). Smaller tiles (e.g., parts of **Financial & Insurance Services**, **Wholesale Trade**, **Arts & Recreation**) indicate niche or lower-employment areas. Altogether, the view sums to about **622,289** jobs, making it easy to see both the dominant sub-sectors and how employment is distributed within each industry.

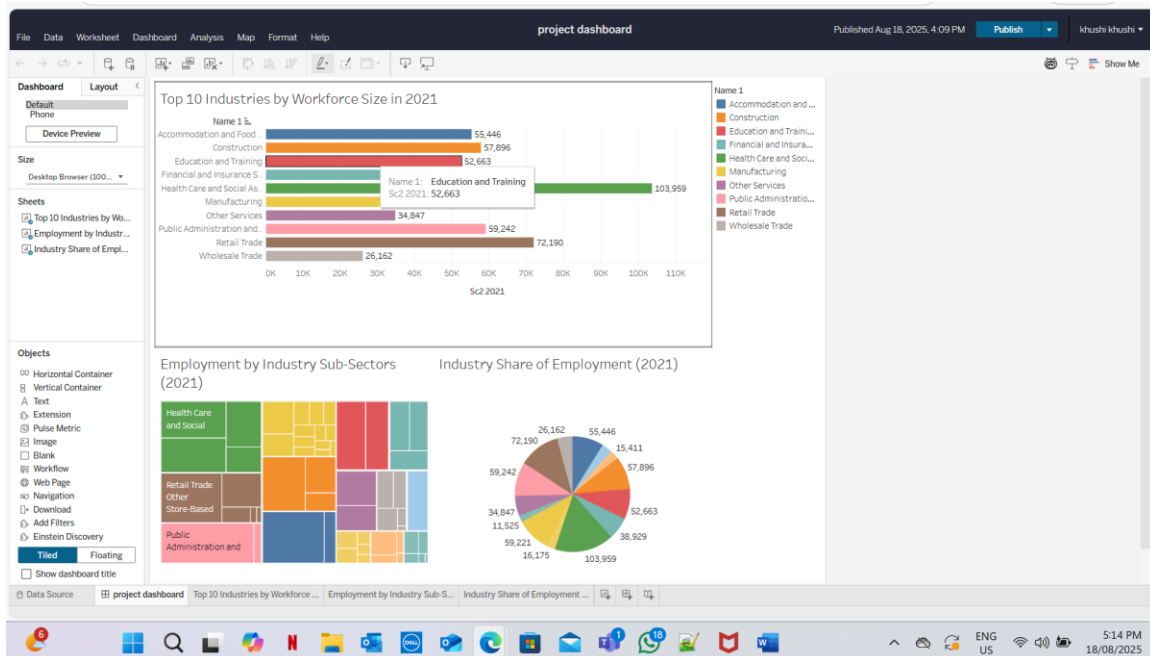


## 16. Data Visualization – Chart 3



This pie chart shows my **industry share of employment in 2021**. Each slice is an industry (Name 1) sized by **SUM(Sc2 2021)**, for a total of **622,289** jobs. **Health Care & Social Assistance** is the largest slice at **103,959 jobs (16.7%)**, followed by **Retail Trade** with **72,190 (11.6%)**. A mid-tier cluster sits around 9–10%: **Public Administration & Safety (59,242; 9.5%)**, **Manufacturing (59,221; 9.5%)**, **Construction (57,896; 9.3%)**, **Accommodation & Food Services (55,446; 8.9%)**, and **Education & Training (52,663; 8.5%)**. Smaller shares include **Financial & Insurance Services (38,929; 6.3%)**, **Other Services (34,847; 5.6%)**, **Wholesale Trade (26,162; 4.2%)**, **Administrative & Support Services (18,623; 3.0%)**, **Information Media & Telecommunications (16,175; 2.6%)**, **Arts & Recreation Services (15,411; 2.5%)**, and **Mining (11,525; 1.9%)**. Overall, the view highlights healthcare as the clear leader, with several industries tightly grouped just below it.

## 17. Dashboard Creation



*This dashboard brings my three views together to tell a clear jobs story for Greater Brisbane in 2021. The top-left bar chart ranks the **top 10 industries by workforce**, showing **Health Care & Social Assistance** as the standout (~104k jobs), followed by **Retail Trade** (~72k) and a tight mid-tier around 55–60k (Construction, Public Administration & Safety, Manufacturing, Accommodation & Food Services, Education & Training). The treemap below-left breaks industries into **sub-sectors**, so I can see where employment concentrates inside each category (e.g., Hospitals, Food & Beverage Services, Construction Services). On the right, the pie chart shows each industry's **share of the total** workforce, reinforcing healthcare's lead and the mid-tier cluster just below it. Together, the three views let me rank industries, drill into sub-sectors, and understand overall market share at a glance.*

### ~Conclusion

By integrating S3, Snowflake, and Tableau, I delivered a scalable, secure, and repeatable workflow that converts raw files into decision-ready insights. The dashboard answers three key questions at a glance—**who employs the most people, where those jobs sit within sub-sectors, and how each industry contributes to the total**—while the Snowflake layer keeps the data model clean and query performance strong for live analytics. If I extend this work, I'll add year parameters and forecasting (2011→2041) for trend exploration, automate ingestion with **Snowpipe**, and introduce region filters to compare SA2/SA3 areas. Overall, the solution meets the assignment's technical requirements and clearly communicates the business story behind Brisbane's employment landscape.

## 18. Project – Hours and Cost Estimation

For this project, I spent a total of **around 14 hours** completing all the required tasks. Setting up Snowflake, creating the database, schema, and tables, and running queries took me about **3 hours**. The **AWS S3 setup**, including bucket creation, data upload, and configuring permissions, required roughly **4 hours**. Writing and executing SQL queries for loading and analyzing the dataset took another **2 hours**. The **Tableau visualizations** and dashboard creation were the most time-consuming, taking around **3 hours**. Finally, preparing the **report and presentation** took about **2 hours**.

### References

- AWS. (2025). *Amazon Simple Storage Service (S3) Documentation*. Amazon Web Services. Retrieved from <https://docs.aws.amazon.com/s3>
- Snowflake Inc. (2025). *Snowflake Documentation*. Retrieved from <https://docs.snowflake.com>
- Tableau Software. (2025). *Tableau Documentation*. Salesforce. Retrieved from <https://www.tableau.com/support/help>
- NIT2202 Group Assignment. (2025). *Live Demonstration Report – Technical Setup (Snowflake, AWS S3, Tableau)*