



GENERAL SIR JOHN KOTELAWALA DEFENCE UNIVERSITY
FACULTY OF MANAGEMENT SOCIAL SCIENCES & HUMANITIES
DEPARTMENT OF LANGUAGES
BACHELOR OF SCIENCE IN APPLIED DATA SCIENCE AND COMMUNICATION
INTAKE 41

LB 2114: ADVANCED SQL AND CLOUD DATABASES
ASSIGNMENT
YEAR 2: SEMESTER 4

GROUP MEMBERS

D/ADC/24/0003 - Thesanya Lamahewa
D/ADC/24/0013 - Samoda De Silva
D/ADC/24/0028 - Thisari Perera
D/ADC/24/0048 - Dilakna Godagamage

Task - 01

Technical Report

Data Ingestion, Cleaning, Analysis and Visualization of UK Managed Migration Datasets

Table of Contents

1. Introduction	4
1.1. Data Gathering and First Impressions.....	5
1.2. Sponsorship data in excel	6
1.3. Create SQL Server	7
1.4. Azure Data Studio	8
1.5. Connect to the database and Server.....	8
2. Creating Views	12
2.1. Database Schema Design and Table Creation	12
2.2. Data Ingestion and the ETL Process	12
2.3. Data Cleaning and Preparation.....	17
2.4. Creating SQL Views for Simplified Reporting and Visualization	19
3. Data Modeling with DAX	19
4. Implementation of Power Bi Dashboard	27
4.1. Dashboard Overview	27
4.2. Key Findings.....	33
5. Challenges During Data Import, Cleaning, and Visualization	34
6. Application of Advanced Power BI Features	34
7. Insights Discovered and Policy Implications	35
8. How SQL Server has improved the way we analyze data with Power BI.....	35
9. Conclusion	36

1. Introduction

The globalization of migration systems has become increasingly complex, and in this environment, there is a growing need to support data-driven analysis to better understand long term trends in international migration flows, as well as evaluate policy responses. In England, recent changes to study and work visa policies have also had a pronounced impact on patterns of international mobility, sponsorship practices and immigration management. In this context, our project uses the Managed Migration Historic datasets, specifically the Study Sponsorship and Work Sponsorship datasets, published by the UK government on <http://data.gov.uk>, to investigate and analyze the most recent applications of this impact. This dataset reports nearly all historical sponsored visas (including approvals, refusals, country of origin, and to what extent the visas are supported). To facilitate the analysis of this dataset, and incorporate it into a SQL Server format, the data is cleaned and modelled into an efficient and usable form to explore large volumes of structured immigration records, the cleaned data is then used in Power BI to connect and visualize and open analysis. This report provides an interactive dashboard to explore long-term trends in visa applications, identifies the leading nationalities in sponsored migration, and looks at outcome patterns including the rates of approval and refusal. The visualization also shows the impacts of policy changes on sponsored migration flows, whether for study or work. In summary, the project highlights the potential to transform government open data into valuable knowledge to support evidence-based examination of migration within the UK.



1.1. Data Gathering and First Impressions

The screenshot shows a web browser displaying a government statistics page. At the top, there's a banner with a circular icon and the text 'up to March 2023.' Below this, a link reads 'For current immigration system data, visit [Immigration system statistics data tables](#)'. The main content area is titled 'Sponsorship'. Under 'Study sponsorship (Confirmation of acceptance for Studies)', it says '(MS Excel Spreadsheet, 1.04 MB)' and describes two datasets: CAS_D01 (confirmation of acceptance for study by institution type) and CAS_D02 (confirmation of acceptance for study by nationality). A note states 'This is not the latest data'. Under 'Work sponsorship (Certificate of Sponsorship)', it says '(MS Excel Spreadsheet, 1.04 MB)' and describes two datasets: CoS_D01 (certificates of sponsorship by industry type) and CoS_D02 (certificates of sponsorship by nationality). Another note states 'This is not the latest data'. At the bottom, there's a link to 'Entry clearance visas granted outside the UK'. A small '↑ Contents' link is at the very bottom left.

Figure 1: Datasets

The first part of the project included obtaining two main datasets from <http://data.gov.uk>,

Managed migration Historic datasets - Sponsorship,

- **Study Sponsorship Data** - This consists of records of visas which are sponsored by educational organizations.
- **Work Sponsorship Data** - This includes records of visas that are sponsored by employers.

1.2. Sponsorship data in excel

Cover_sheet

Home Office

NATIONAL STATISTICS

Immigration System Statistics

3 year ending March 2023

4 Study sponsorship (Confirmation of Acceptance for Studies)

5 Published: 25 May 2023

6 Next update: 24 August 2023

7 Responsible Statistician: Rex Newell

8 Email: MigrationStatsEnquiries@homeoffice.gov.uk

9 Press enquiries: 0300 123 3535

10 Crown copyright © 2023

List of Fields

Field	Description
2 Year / quarter	The period in which the Confirmation was used in an application.
3 Type of application	Whether the CAS was used in an entry-clearance visa application (visa), or an application for extension of leave in the UK (extension).
4 Institution type group	The type of institution (grouped) for which the application to study was made (Higher education, other).
5 Institution type	The type of institution for which the application to study was made.
6 Geographical region	World region of the country of nationality for applicant.
7 Nationality	Nationality of the applicant.
8 Applications	The total number of applications which used a CAS in either a visa or extension of stay application.

migration-study-sponsorship-datasets-mar-2023

Search for tools, help, and more (Alt + Q)

File Home Insert Share Page Layout Formulas Data Review View Help Draw

Applications for visas and extensions of stay for study using sponsor acceptances, by education sector, 2010 to 2023 Q1

Year	Column Labels	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4	2023 Q1
6 Visa	20,872	228,325	63,636	58,232	37,659	303,133	85,334	51,264
7 Higher Education institution	16,740	209,109	56,455	52,711	31,674	277,990	77,083	46,355
8 Russell Group Universities	6,101	104,456	15,242	6,174	8,702	116,355	8,614	3,528
9 Non Russell Group Universities	10,639	104,653	41,213	46,537	22,972	161,635	68,469	42,827
10 All other institutions	4,132	19,216	7,181	5,521	5,985	25,143	8,251	4,909
11 Tertiary, Further education or other colleges	882	5,872	2,801	2,006	1,727	9,591	3,668	2,160
12 Independent school	2,492	8,929	1,370	1,016	3,028	8,954	1,395	1,088
13 English language school	371	2,997	2,550	2,230	438	4,854	2,698	1,235
14 Other	387	1,418	460	269	792	1,744	490	426
15 Extensions of stay	3,423	9,237	12,542	6,202	2,731	13,588	20,958	6,416
16 Higher Education institution	3,155	8,368	11,570	5,687	2,455	12,293	19,446	6,147
17 Russell Group Universities	788	4,071	5,689	1,879	887	7,763	11,613	2,395
18 Non Russell Group Universities	2,367	4,297	5,881	3,808	1,568	4,530	7,833	3,752
19 All other institutions	268	868	972	515	276	1,295	1,512	269
20 Tertiary, Further education or other colleges	51	262	266	197	38	335	423	75
21 Independent school	180	523	617	165	181	836	945	149
22 English language school	14	15	44	101	26	77	99	31

Figure 2: Dataset in Excel

The first examination of the Excel screenshots suggests the data is collected into a semi structured wide table format that is common to government reported statistics. The tables are appropriate for reporting summaries but will require substantial reshaping to support well-structured relational data storage and analysis.

1.3. Create SQL Server

Through the Azure Portal, we provisioned the Azure SQL Database environment; we created a new logical server, which we named ukimmigration server, and then created the store database, UKImmigrationDB, within that logical server and selected a cost-effective General-Purpose Serverless Compute Tier for the deployment. The deployment was completed successfully, which established the starting point of the cloud data foundation for this project

Create SQL Database

Subscription: Azure for Students

Resource group: KDU

Database name: UKImmigrationDB

Server: (new) ukimmigration-server (Southeast Asia)

Want to use SQL elastic pool? No

Workload environment: Development

Microsoft.SQLDatabase.newDatabaseNewServer_1f6bc7a80986482a99ec6 | Overview

Your deployment is complete

Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_1f6bc7a80986482a99ec6

Subscription: Azure for Students

Resource group: KDU

Start time: 15/11/2025, 13:15:20

Correlation ID: 7f5f82b-2545-489c-8b45-210be4f02bae

Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials
[Start learning today >](#)

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

Figure 3: Create SQL Database

1.4. Azure Data Studio

Azure Data Studio interface was set up to connect to an instance of SQL Server securely, which included connection parameters for the instance, which was likely local host if local or some server if cloud, and could be either Windows or SQL Server authentication.

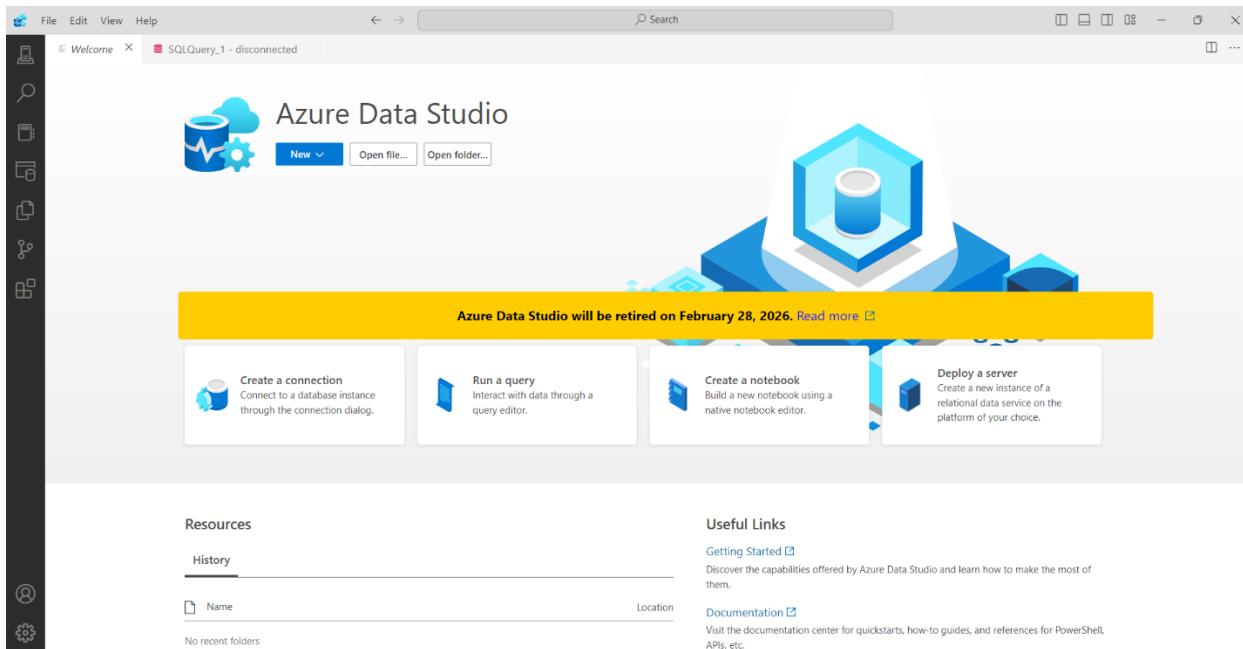


Figure 4: Azure Data Studio

1.5. Connect to the database and Server

After the successful provisioning of the Azure SQL server and database, a connection was established to the ukimmigration server from within Azure Data Studio. This interface was highly useful in writing, executing, and debugging the complex SQL scripts needed for this project. Using Azure Data Studio, we imported two key UK Managed Migration datasets: Study Sponsorship and Work Sponsorship. We started cleaning the data. SQL files kept these scripts organized for reproducibility. Azure Data Studio was instrumental throughout the data engineering lifecycle that started with the initial data ingestion and cleaning to exploratory data analysis.

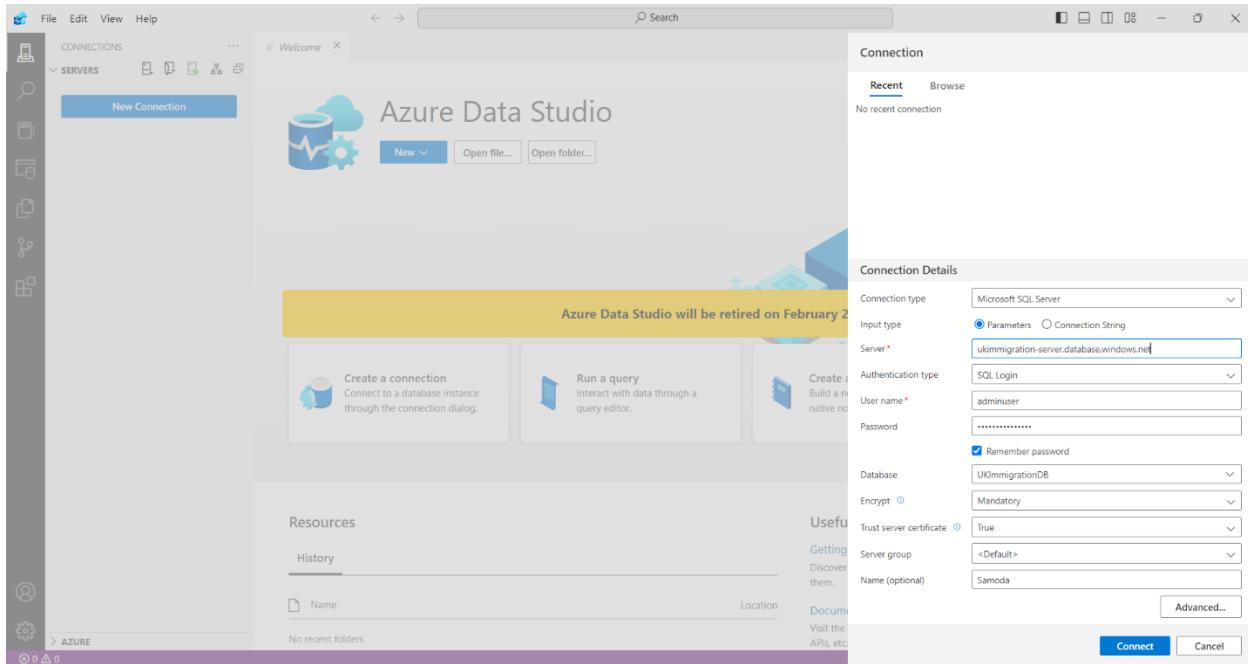


Figure 5: Connect SQL

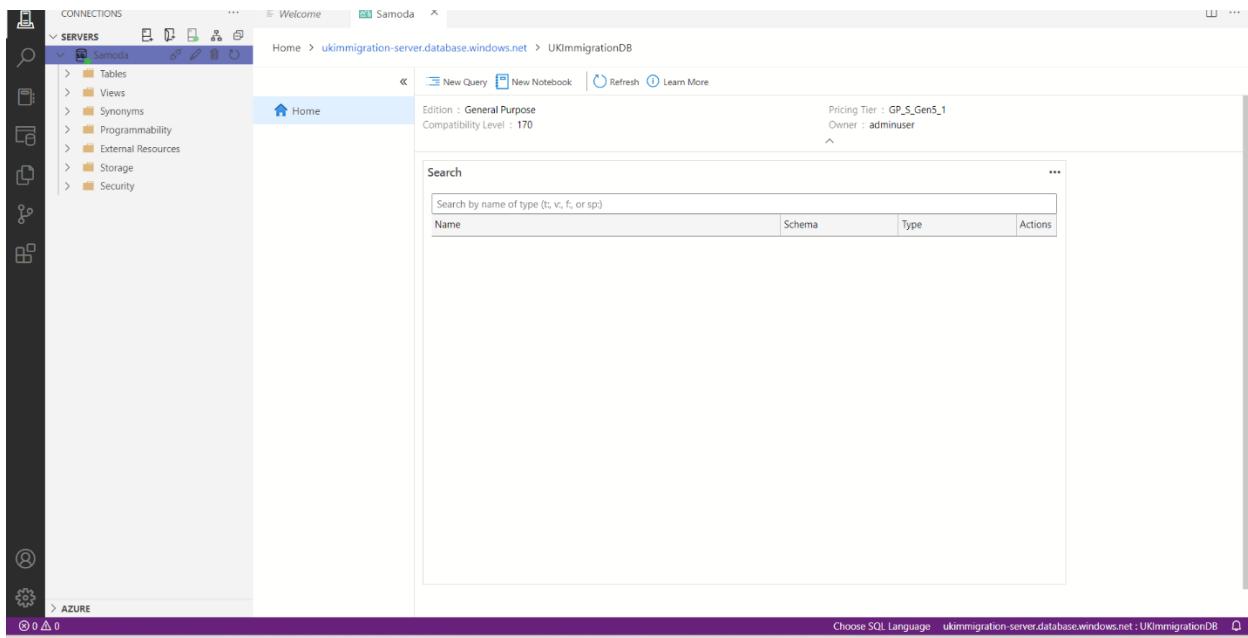


Figure 6: Creating Database

firewall rule addition

The addition of firewall rules to the Azure Portal allows for secure communication between the Azure Data Studio and Cloud SQL Server via Firewall Rules. Only IP addresses defined by the user in their Firewall rule can connect to their database. Firewall rules allow owners to control which IP addresses connect directly to their database and provide strong security against unauthorized access to data.

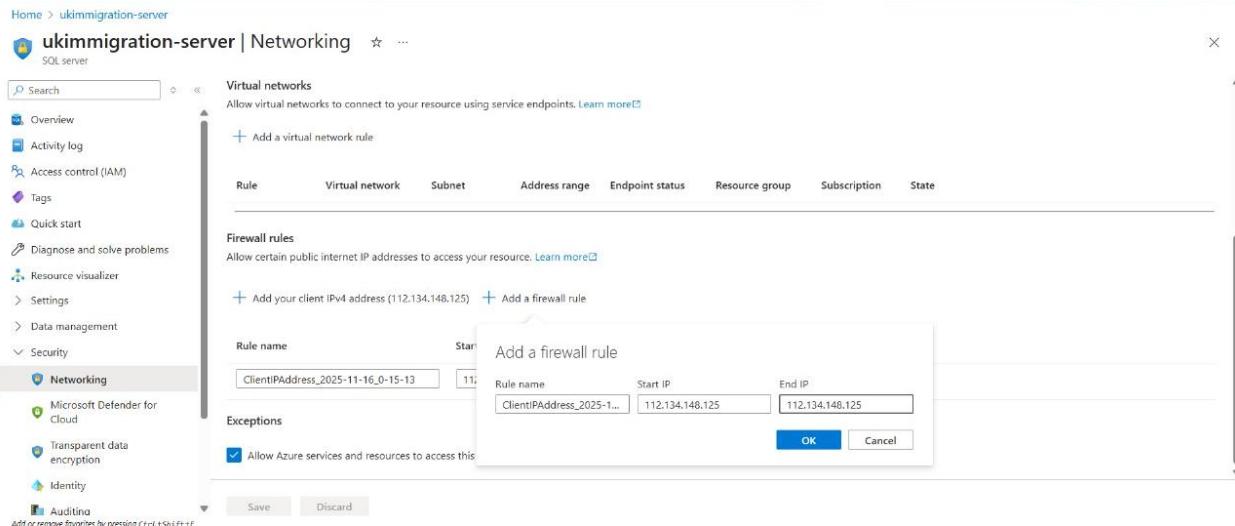


Figure 7: Add Firewall Rule

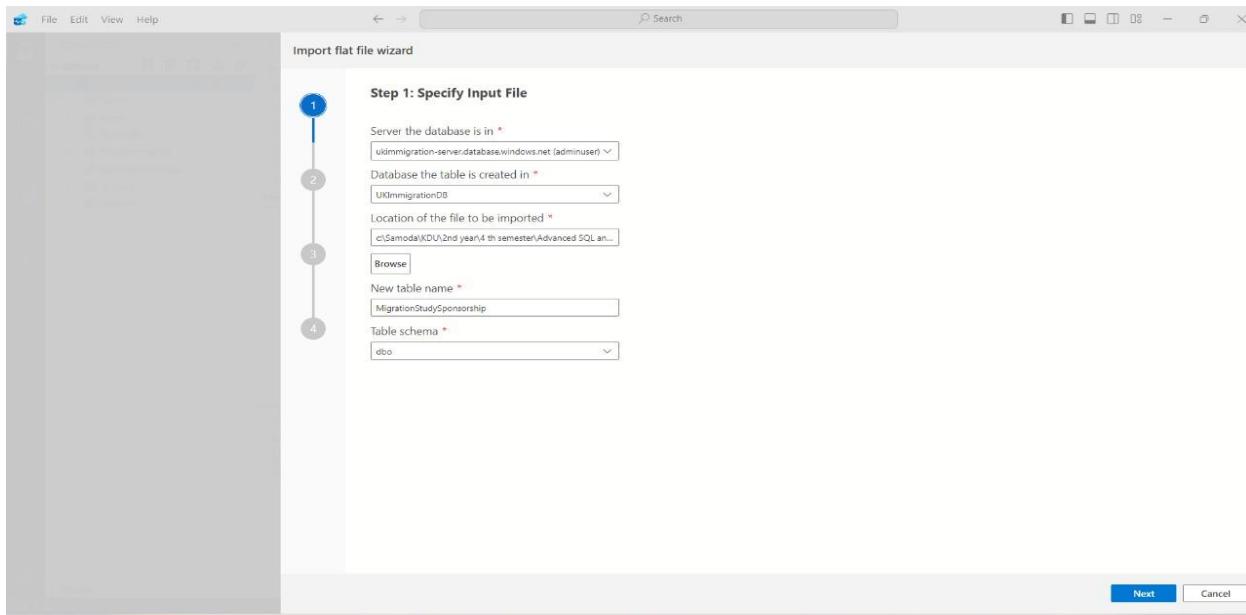


Figure 8: Step 01

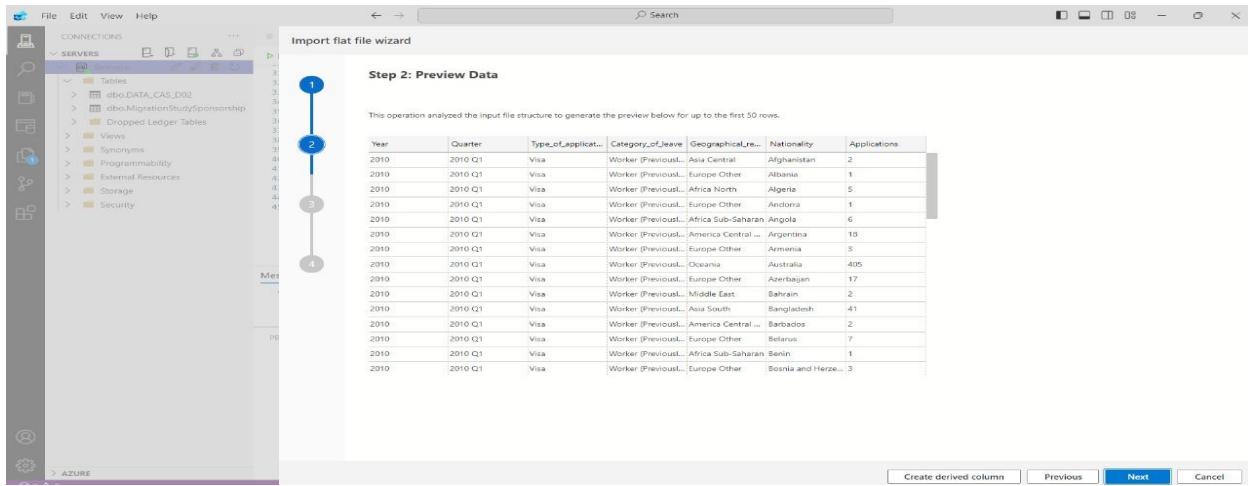


Figure 9: Step 02

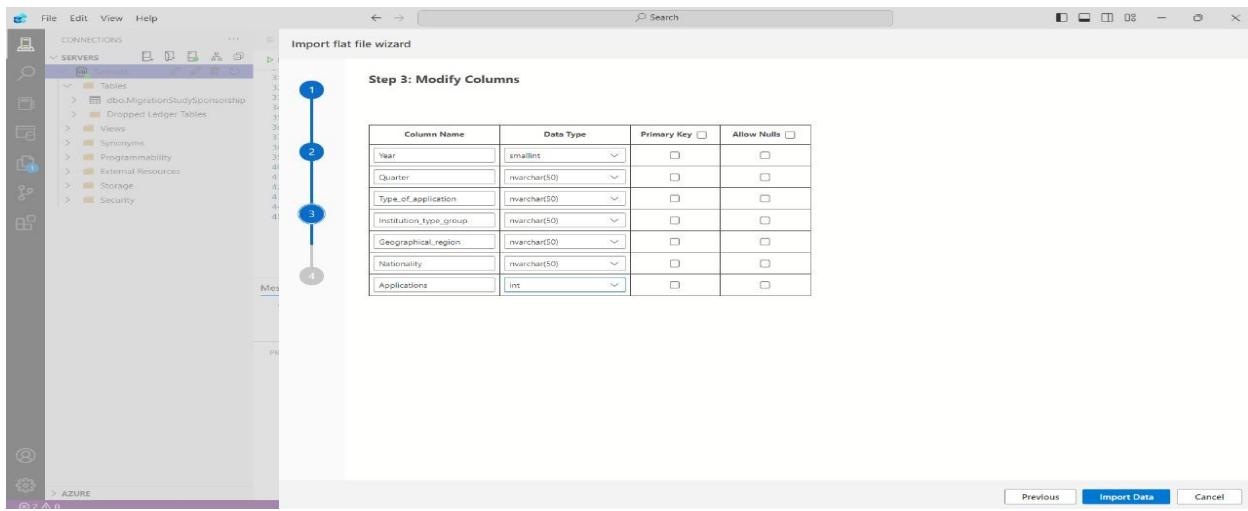


Figure 10: Step 03

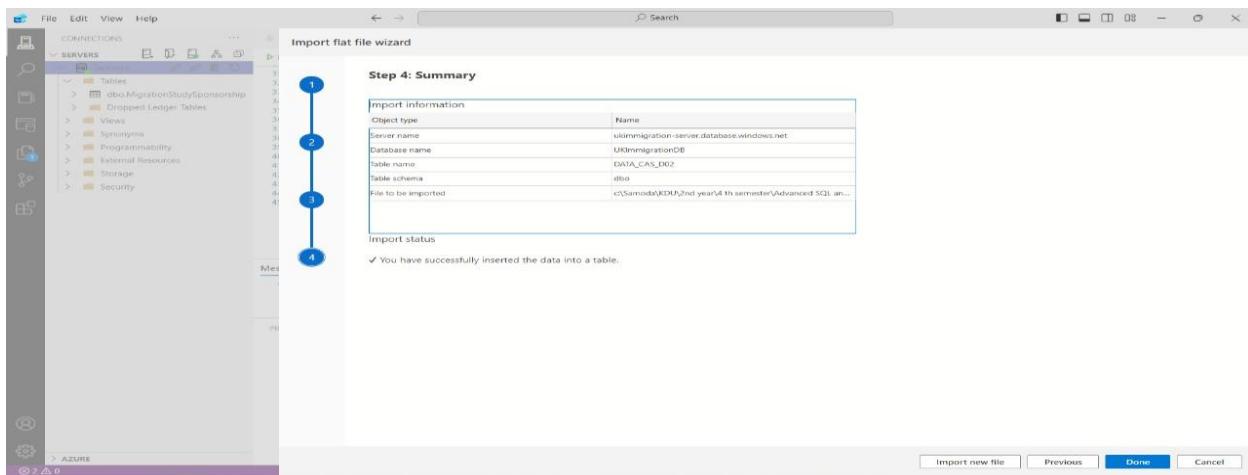


Figure 11: Step 04

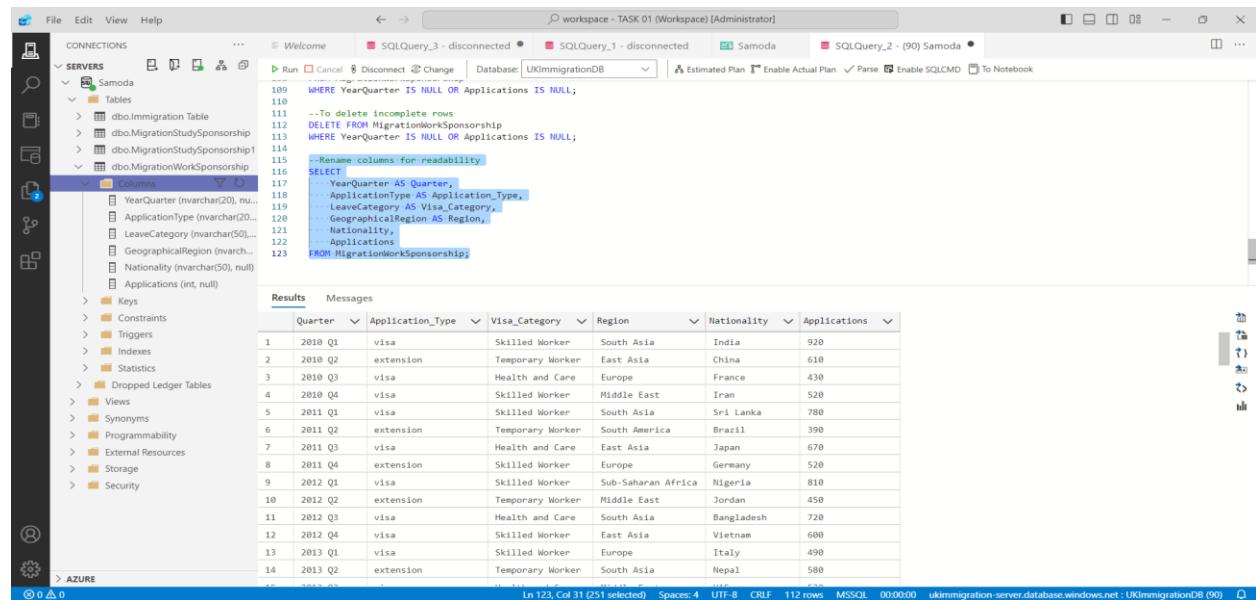
2. Creating Views

2.1. Database Schema Design and Table Creation

The database schema was designed to structure the raw Excel data. The scheme contains the two major tables for this project, such as Study Sponsorship and Work Sponsorship. The CREATE TABLE statements were written from within Azure Data Studio to create the physical tables, using different columns with their corresponding data types (VARCHAR for strings, INT for integer count values) to help maintain data quality and to provide for optimal storage conditions. To ensure data integrity, it is necessary that some columns have been declared as NOT NULL. Key columns include Year and Sponsor Name. All data analysis would be based on these columns.

2.2. Data Ingestion and the ETL Process

To incorporate the original data from Excel into SQL Server, we made use of ETL tools, meaning Extract, Transform, and then finally Load data from Excel into SQL Server. Transformation was performed both at the time of importing into SQL Server (import), and via custom SQL Scripts. Transformation consists of data profiling activities that expose and remove problems related to formatting, such as mixed formats or missing data, prior to loading cleaned data into the Study Sponsorship and Work Sponsorship Tables. Therefore, the result of ETL processing was a dynamic, quarriable database asset in a DBMS (Database Management System) that could be used as needed.



A screenshot of the Azure Data Studio interface. The left sidebar shows a tree view of 'Samoda' database objects, including 'Tables' which contain 'dbo.MigrationTable', 'dbo.MigrationStudySponsorship', 'dbo.MigrationWorkSponsorship', and 'dbo.UKImmigrationDB'. The main pane displays a SQL query window with the following code:

```
100 WHERE YearQuarter IS NULL OR Applications IS NULL;
101
102 --To delete incomplete rows
103 DELETE FROM MigrationWorkSponsorship
104 WHERE YearQuarter IS NULL OR Applications IS NULL;
105
106 --Rename columns for readability
107 SELECT
108     YearQuarter AS Quarter,
109     ApplicationType AS Application_Type,
110     LeaveCategory AS Visa_Category,
111     GeographicalRegion AS Region,
112     Nationality,
113     Applications
114     FROM MigrationWorkSponsorship;
```

The results pane shows a table with 14 rows of data, mapping the renamed columns to their original names:

	Quarter	Application_Type	Visa_Category	Region	Nationality	Applications
1	2010 Q1	visa	Skilled Worker	South Asia	India	920
2	2010 Q2	extension	Temporary Worker	East Asia	China	610
3	2010 Q3	visa	Health and Care	Europe	France	430
4	2010 Q4	visa	Skilled Worker	Middle East	Iran	520
5	2011 Q1	visa	Skilled Worker	South Asia	Sri Lanka	780
6	2011 Q2	extension	Temporary Worker	South America	Brazil	390
7	2011 Q3	visa	Health and Care	East Asia	Japan	670
8	2011 Q4	extension	Skilled Worker	Europe	Germany	520
9	2012 Q1	visa	Skilled Worker	Sub-Saharan Africa	Nigeria	810
10	2012 Q2	extension	Temporary Worker	Middle East	Jordan	450
11	2012 Q3	visa	Health and Care	South Asia	Bangladesh	720
12	2012 Q4	visa	Skilled Worker	East Asia	Vietnam	600
13	2013 Q1	visa	Skilled Worker	Europe	Italy	490
14	2013 Q2	extension	Temporary Worker	South Asia	Nepal	580

Figure 12: Rename Columns

```

-- Rename columns for readability
SELECT
    YearQuarter AS Quarter,
    ApplicationType AS Application_Type,
    LeaveCategory AS Visa_Category,
    GeographicalRegion AS Region,
    Nationality,
    Applications
FROM MigrationWorkSponsorship;

-- Exploratory Data Analysis
-- Applications over time
SELECT YearQuarter, SUM(Applications) AS TotalApplications
FROM MigrationWorkSponsorship
GROUP BY YearQuarter
ORDER BY YearQuarter;

```

YearQuarter	TotalApplications
1 2010 Q1	1848
2 2010 Q2	1220
3 2010 Q3	868
4 2010 Q4	1048
5 2011 Q1	1568
6 2011 Q2	788
7 2011 Q3	1340
8 2011 Q4	1048
9 2012 Q1	1628
10 2012 Q2	908
11 2012 Q3	1448
12 2012 Q4	1280
13 2013 Q1	988
14 2013 Q2	1168
-- 2013 Q3	1088

Figure 13: Exploratory Data Analysis

```

-- Rename columns permanently:
EXEC sp_rename 'MigrationStudySponsorship.TypeOfApplication', 'Application_Type', 'COLUMN';
EXEC sp_rename 'MigrationStudySponsorship.InstitutionTypeGroup', 'Institution_Group', 'COLUMN';
EXEC sp_rename 'MigrationStudySponsorship.GeographicalRegion', 'Region', 'COLUMN';

-- Exploratory Data Analysis (EDA)
-- 1. Applications Over Time
SELECT YearQuarter, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY YearQuarter
ORDER BY YearQuarter;

-- 2. Applications by Region
SELECT Region, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY Region
ORDER BY TotalApplications DESC;

```

YearQuarter	TotalApplications
1 2010 Q1	958
2 2010 Q2	628
3 2010 Q3	438
4 2010 Q4	518
5 2011 Q1	788
6 2011 Q2	398
7 2011 Q3	678
8 2011 Q4	528
9 2012 Q1	818
10 2012 Q2	458
11 2012 Q3	728

Figure 14: Applications Over Time

The screenshot shows the SQL Server Management Studio interface. On the left, the Object Explorer displays a database named 'Samoda' with tables like 'dbo.Immigration Table', 'dbo.MigrationStudySponsorship', and 'dbo.MigrationWorkSponsorship'. The 'dbo.MigrationWorkSponsorship' table is selected, and its 'Columns' folder is expanded, showing columns such as 'YearQuarter', 'LeaveCategory', 'GeographicalRegion', 'Nationality', and 'Applications'. The main pane contains a SQL query window with the following code:

```

124 --Exploratory Data Analysis
125 --Applications over time
126 SELECT YearQuarter, SUM(Applications) AS TotalApplications
127 FROM MigrationWorkSponsorship
128 GROUP BY YearQuarter
129 ORDER BY YearQuarter;
130
131
132 --Applications by region
133 SELECT GeographicalRegion, SUM(Applications) AS TotalApplications
134 FROM MigrationWorkSponsorship
135 GROUP BY GeographicalRegion
136 ORDER BY TotalApplications DESC;
137
138
139 --Applications by Leave Category
140 SELECT LeaveCategory, SUM(Applications) AS TotalApplications
141 FROM MigrationWorkSponsorship
142 GROUP BY LeaveCategory;

```

The results pane shows a table with the following data:

GeographicalRegion	TotalApplications
South Asia	25900
East Asia	14240
Europe	10780
Middle East	7580
Sub-Saharan Africa	6660
South America	5400

Figure 15: Application by region

The screenshot shows the SQL Server Management Studio interface. The left pane shows the same database structure as Figure 15. The 'dbo.MigrationWorkSponsorship' table is selected, and its 'Columns' folder is expanded. The main pane contains a SQL query window with the following code:

```

128 FROM MigrationWorkSponsorship
129 GROUP BY YearQuarter;
130 ORDER BY YearQuarter;
131
132
133 SELECT GeographicalRegion, SUM(Applications) AS TotalApplications
134 FROM MigrationWorkSponsorship
135 GROUP BY GeographicalRegion
136 ORDER BY TotalApplications DESC;
137
138
139 --Applications by Leave Category
140 SELECT LeaveCategory, SUM(Applications) AS TotalApplications
141 FROM MigrationWorkSponsorship
142 GROUP BY LeaveCategory;

```

The results pane shows a table with the following data:

LeaveCategory	TotalApplications
Health and Care	16100
Skilled Worker	41000
Temporary Worker	13460

Figure 16: Application by Leave Category

The screenshot shows the SSMS interface with the following details:

- File Bar:** File, Edit, View, Help
- Toolbar:** Welcome, Run, Cancel, Disconnect, Change, Database: UKImmigrationDB, Estimated Plan, Enable Actual Plan, Parse, Enable SQLCMD, To Notebook.
- Servers Tree:** SERVERS > Samoda > Tables > dbo.Immigration Table, dbo.MigrationStudySponsorship, dbo.MigrationWorkSponsorship > Columns > Nationality, Applications.
- Query Editor:** Contains the following T-SQL code:

```

134     FROM MigrationWorkSponsorship
135     GROUP BY GeographicalRegion
136     ORDER BY TotalApplications DESC;
137
138     --Applications by Leave Category
139     SELECT LeaveCategory, SUM(Applications) AS TotalApplications
140     FROM MigrationWorkSponsorship
141     GROUP BY LeaveCategory;
142
143     -- Top 10 Nationalities
144     SELECT Nationality, SUM(Applications) AS TotalApplications
145     FROM MigrationWorkSponsorship
146     GROUP BY Nationality
147     ORDER BY TotalApplications DESC
148     OFFSET 0 ROWS FETCH NEXT 10 ROWS ONLY;
149
150

```
- Results Grid:** Displays the top 10 nationalities with their total applications:

Nationality	TotalApplications
India	10640
Sri Lanka	6040
Pakistan	3600
Bangladesh	3020
China	2960
Japan	2700
Nepal	2580
Italy	2000
Germany	1880
Nigeria	1620
- Status Bar:** In 150, Col 1 (213 selected), Spaces: 4, UTF-8, CRLF, 10 rows, MSSQL, 00:00:00, ukimmigration-server.database.windows.net : UKImmigrationDB (90)

Figure 17: Top 10 Nationalities

The screenshot shows the SSMS interface with the following details:

- File Bar:** File, Edit, View, Help
- Toolbar:** Run, Cancel, Disconnect, Change, Database: UKImmigrationDB, Estimated Plan, Enable Actual Plan, Parse, Enable SQLCMD, To Notebook.
- Servers Tree:** SERVERS > Samoda > Tables > dbo.Immigration Table, dbo.MigrationStudySponsorship > Columns > Region, Applications.
- Query Editor:** Contains the following T-SQL code:

```

140     ORDER BY YearQuarter;
141
142     --2. Applications by Region
143     SELECT Region, SUM(Applications) AS TotalApplications
144     FROM MigrationStudySponsorship
145     GROUP BY Region
146     ORDER BY TotalApplications DESC;
147
148     --3. Applications by Institution Group
149     SELECT Institution_Group, SUM(Applications) AS TotalApplications
150     FROM MigrationStudySponsorship
151     GROUP BY Institution_Group;
152
153     --4. Top 10 Nationalities by Application Volume
154     SELECT Nationality, SUM(Applications) AS TotalApplications
155     FROM MigrationStudySponsorship
156     GROUP BY Nationality
157     ORDER BY TotalApplications DESC
158     OFFSET 0 ROWS FETCH NEXT 10 ROWS ONLY;

```
- Results Grid:** Displays the top 6 regions with their total applications:

Region	TotalApplications
South Asia	12980
East Asia	7130
Europe	5390
Middle East	3780
Sub-Saharan Africa	3330
South America	2700
- Status Bar:** Loading ...

Figure 18: Applications By Region

```

--3. Applications by Institution Group
SELECT Institution_Group, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY Institution_Group;

--4. Top 10 Nationalities by Application Volume
SELECT Nationality, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY Nationality
ORDER BY TotalApplications DESC
OFFSET 0 ROWS FETCH NEXT 10 ROWS ONLY;

```

Institution_Group	TotalApplications
Higher education	25730
Other	9580

Figure 19: Application By institution Group

```

--3. Applications by Institution Group
SELECT Institution_Group, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY Institution_Group;

--4. Top 10 Nationalities by Application Volume
SELECT Nationality, SUM(Applications) AS TotalApplications
FROM MigrationStudySponsorship
GROUP BY Nationality
ORDER BY TotalApplications DESC
OFFSET 0 ROWS FETCH NEXT 10 ROWS ONLY;

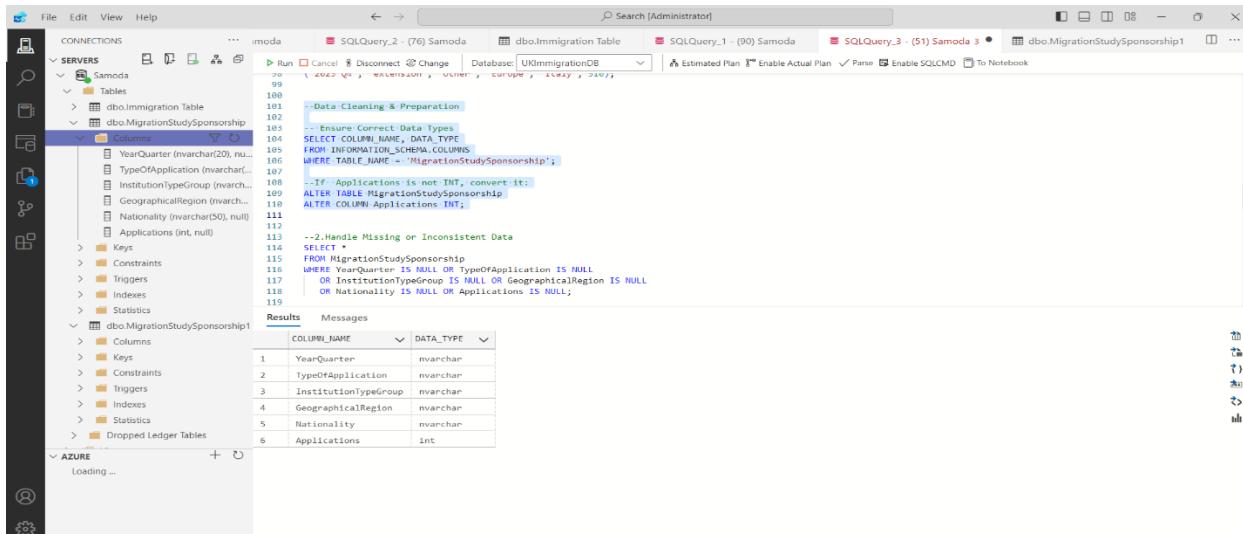
```

Nationality	TotalApplications
India	5050
Sri Lanka	3020
Pakistan	1800
Bangladesh	1510
China	1490
Japan	1350
Nepal	1290
Italy	1000
Germany	940
Nigeria	810

Figure 20: Top Nationalities by Application Volume

2.3. Data Cleaning and Preparation

After ingesting the data, using the SQL scripts included in Azure Data Studio, we were able to conduct an extensive amount of data cleansing, including making references to country names uniform (spelling), replacing empty values with either a zero or Unknown, and validating the numerical columns of data. This process was very important to remove the inconsistencies and errors contained in the original datasets as it provided a clean and accurate base for conducting all future analyses and visualizations.



The screenshot shows the Azure Data Studio interface with a database connection to 'Samoda'. In the center, a SQL query editor window displays a script for data cleaning. The script includes code to ensure correct data types for columns like 'YearQuarter' and 'Applications', and to handle missing or inconsistent data by selecting rows where specific columns are NULL. Below the script, the 'Results' tab shows a table with the results of the SELECT query, listing columns 'YearQuarter', 'TypeOfApplication', 'InstitutionTypeGroup', 'GeographicalRegion', 'Nationality', and 'Applications' along with their data types.

COLUMN_NAME	DATA_TYPE
1 YearQuarter	nvarchar
2 TypeOfApplication	nvarchar
3 InstitutionTypeGroup	nvarchar
4 GeographicalRegion	nvarchar
5 Nationality	nvarchar
6 Applications	int

Figure 21: Data Cleaning

Processes through which there will be an assurance of the quality and uniformity of data through utilization of SQL scripts. The primary methods utilized to carry out these SQL scripts were to utilize SQL statements to modify column data types within the existing database tables by utilizing an ALTER TABLE statement, and also performing SELECT statements with a WHERE clause to identify and rectify NULL values within the data, along with any values within the database that are inconsistent with each other to provide a reliable basis for data analysis.

The screenshot shows the SSMS interface with the following details:

- File, Edit, View, Help** menu bar.
- CONNECTIONS** pane on the left.
- Servers** node expanded, showing **Samoda** and its **Tables**.
- dbo.MigrationStudySponsorship** table selected.
- Columns** node expanded under the table.
- Script** pane showing T-SQL code:

```
ALTER TABLE MigrationStudySponsorship
ALTER COLUMN Application INT;
--2. Handle Missing or Inconsistent Data
SELECT *
FROM MigrationStudySponsorship
WHERE YearQuarter IS NULL OR TypeOfApplication IS NULL
OR InstitutionTypeGroup IS NULL OR GeographicalRegion IS NULL
OR Nationality IS NULL OR Applications IS NULL;
--To remove incomplete rows:
DELETE FROM MigrationStudySponsorship
WHERE YearQuarter IS NULL OR Applications IS NULL;
--3. Rename Columns for Readability
SELECT
    YearQuarter AS Quarter,
```

- Results** tab selected, showing the schema of the table:

YearQuarter	TypeOfApplication	InstitutionTypeGroup	GeographicalRegion	Nationality	Applications
-------------	-------------------	----------------------	--------------------	-------------	--------------

Figure 22: Handling Missing and Incomplete Data

There is an established link between Azure Data Studio and the ukimmigration server based on the use of Object Explorer, which ensures the correctness of the database schema described in the publication, including the names of the two main tables, dbo.MigrationStudySponsorship, which is populated by two different sets of data, and dbo.immigrationTable, which includes all the required fields with respective data types.

The screenshot shows the SSMS interface with the following details:

- File, Edit, View, Help** menu bar.
- CONNECTIONS** pane on the left.
- Servers** pane showing a connection to **Samoda**.
- Tables** node under **dbo** expanded, showing **dbo.MigrationStudySponsorship**.
- Columns** node under **dbo.MigrationStudySponsorship** is selected.
- Script** pane showing T-SQL code for column renaming:

```
--3. Rename Columns for Readability
SELECT
    YearQuarter AS Quarter,
    TypeOfApplication AS Application_Type,
    InstitutionTypeGroup AS Institution_Group,
    GeographicalRegion AS Region,
    Nationality,
    Applications
FROM MigrationStudySponsorship;
--Rename columns permanently
EXEC sp_rename 'MigrationStudySponsorship.TypeOfApplication', 'Application_Type', 'COLUMN';
EXEC sp_rename 'MigrationStudySponsorship.InstitutionTypeGroup', 'Institution_Group', 'COLUMN';
EXEC sp_rename 'MigrationStudySponsorship.GeographicalRegion', 'Region', 'COLUMN';
--Exploratory Data Analysis (EDA)
--1. Applications Over Time
SELECT YearQuarter, SUM(Applications) AS TotalApplications
```

- Results** pane showing a table with 12 rows of data:

	Quarter	Application_Type	Institution_Group	Region	Nationality	Applications
1	2010 Q1	visa	Higher education	South Asia	India	950
2	2010 Q2	extension	Other	East Asia	China	620
3	2010 Q3	visa	Higher education	Europe	France	430
4	2010 Q4	visa	Higher education	Middle East	Iran	510
5	2011 Q1	visa	Higher education	South Asia	Sri Lanka	780
6	2011 Q2	extension	Other	South America	Brazil	390
7	2011 Q3	visa	Higher education	East Asia	Japan	670
8	2011 Q4	extension	Other	Europe	Germany	520
9	2012 Q1	visa	Higher education	Sub-Saharan Africa	Nigeria	810
10	2012 Q2	extension	Other	Middle East	Jordan	450
11	2012 Q3	visa	Higher education	South Asia	Bangladesh	720
12	2012 Q4	visa	Higher education	East Asia	Vietnam	680

Figure 23: Rename Columns for Readability

2.4. Creating SQL Views for Simplified Reporting and Visualization

Developing SQL Views for Easier Reporting and Visualization In order to facilitate the reporting workflow between the SQL database and a reporting tool such as Power BI, we developed several SQL Views. A View is a virtual table defined by a stored query which can abstract some of the more profitable logic and allow for a simpler and focused size of data. For instance, we could have developed a view named vw_WorkVisa_ApprovalRates which is based on a query that joins relevant tables and calculates our approval rate percentage ahead of time. By using Azure Data Studio to develop Views, we were able to introduce an abstraction that enables the user to interact with the data without requiring an understanding of the complexities associated with the schema on which the Power BI Report is based. By decoupling the process of developing reports and schema, we were able to make developing reports within Power BI simpler and more error-free. At the same time, it also leads towards reusability and retains a single source of meaningful data contained within the SQL database.

3. Data Modeling with DAX

We have set up various DAX formulas that make it easier for us to calculate these advanced metrics, which are available for review within measures created for approval rates, refusal rates, and Year Over Year percent changes. DAX is an interactive and dynamic language that enables us to calculate all the metrics that we can review or calculate based on user interactions, for instance, based on a year, country, or sponsor type. DAX plays an integral role within interactive reporting.

The screenshot shows the Microsoft Power BI Data Editor interface. The top navigation bar includes File, Home, Help, Table tools, and Measure tools. The main area displays a table with columns: Year, Quarter, Application_Type, Visa_Category, Industry_Sector, and Applications. The table data is as follows:

Year	Quarter	Application_Type	Visa_Category	Industry_Sector	Applications
2010	2010 Q1	Visa	Worker (Previously Tier 2)	Electricity, gas, steam and air conditioning supply	32
2010	2010 Q1	Extensions of stay	Worker (Previously Tier 2)	Accommodation and Food Service Activities	420
2010	2010 Q2	Visa	Worker (Previously Tier 2)	Mining and Quarrying	144
2010	2010 Q2	Extensions of stay	Worker (Previously Tier 2)	Agriculture, Forestry and Fishing	11
2010	2010 Q3	Extensions of stay	Temporary Worker (Previously Tier 5)	Professional, Scientific and Technical Activities	2
2010	2010 Q4	Visa	Worker (Previously Tier 2)	Accommodation and Food Service Activities	318
2010	2010 Q4	Extensions of stay	Worker (Previously Tier 2)	Mining and Quarrying	54
2019	2019 Q4	Unknown	Worker (Previously Tier 2)	Activities of extraterritorial organisations and bodies	4

The formula bar at the top shows the DAX code for the 'Total Applications' measure:

```
1 Total Applications by Application TypeTotalApplicationsByType =  
2 CALCULATE(  
3 COUNTA([Applications]),  
4 ALLEXCEPT(  
5 'DATA_CAS_D01 (Work)',  
6 'DATA_CAS_D01 (Work)'[Application_Type],  
7 'DATA_CAS_D01 (Work)'[Quarter]  
8 )  
9 )  
10
```

The right side of the screen shows the Power BI Data view pane with the data model structure.

Figure 24: Core Performance Measures

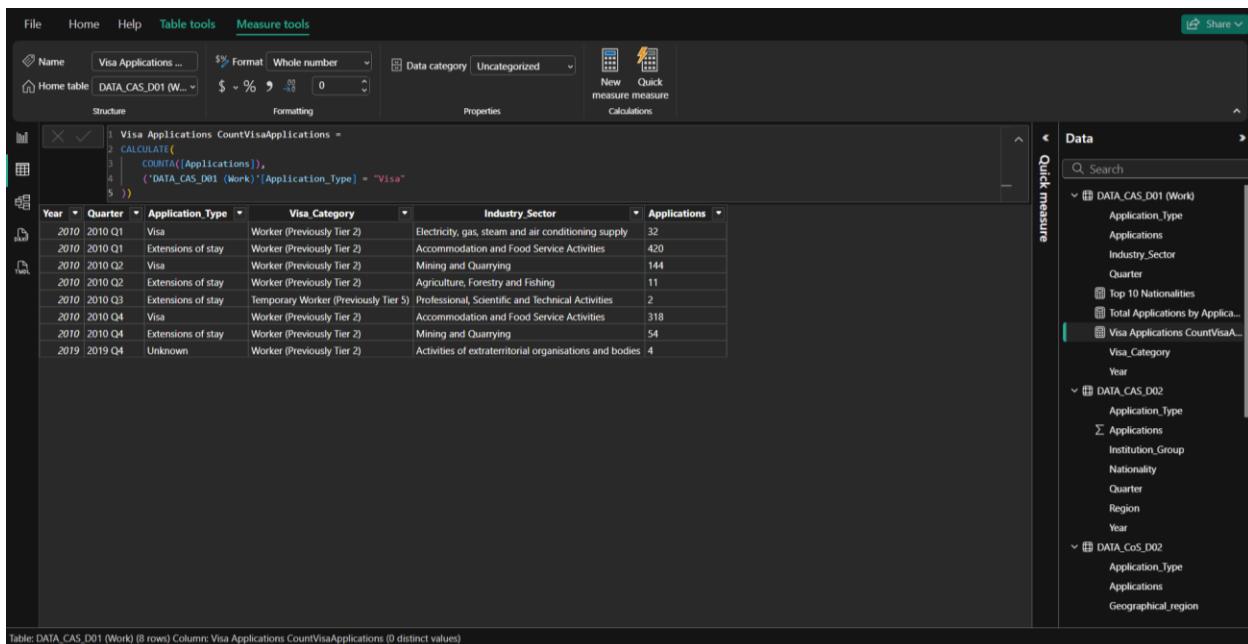


Figure 25: Visa Applications Count

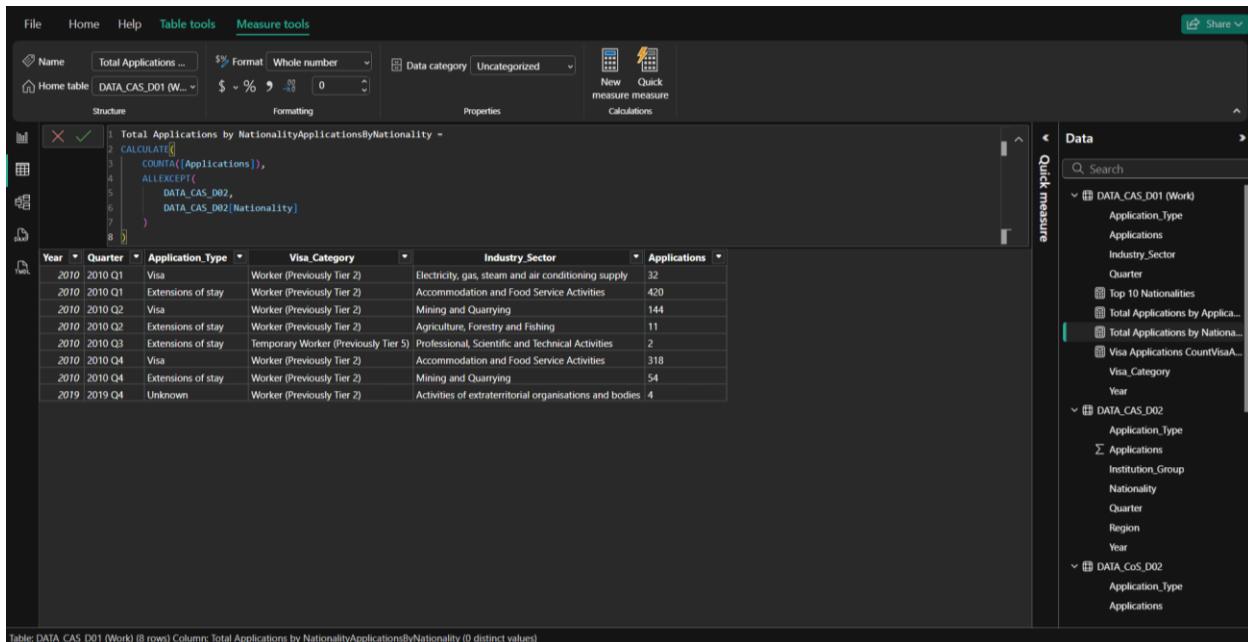


Figure 26: Total Applications by Nationality

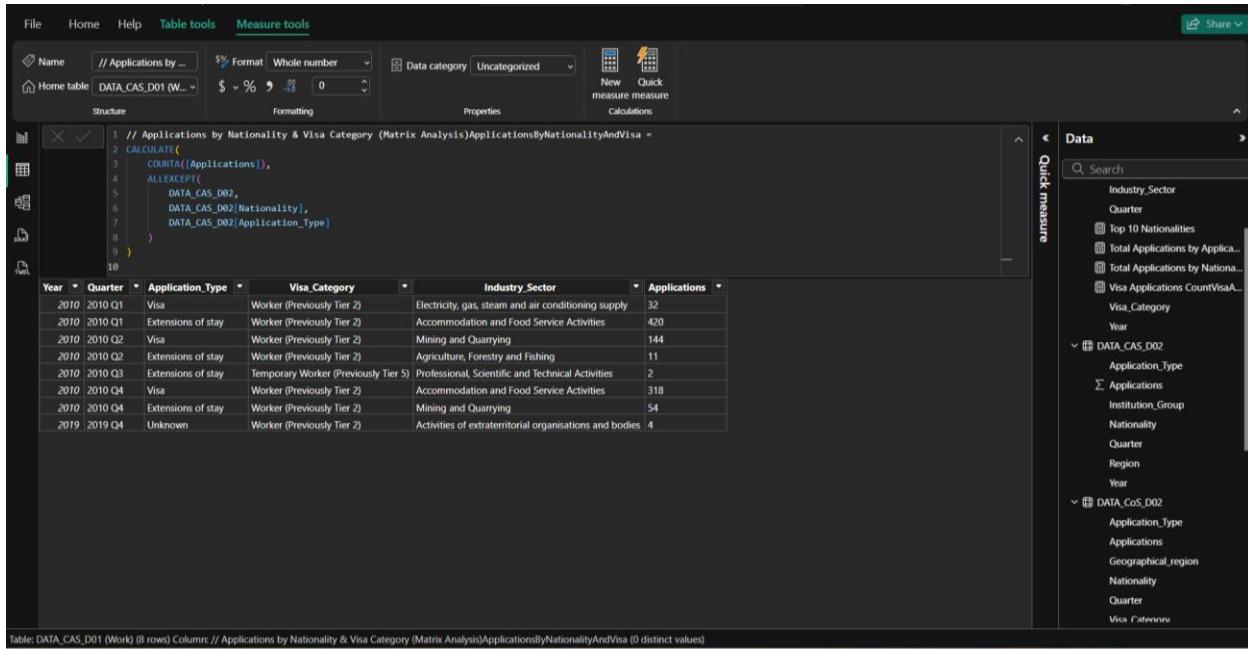


Figure 27: Applications By Nationality and Visa Category

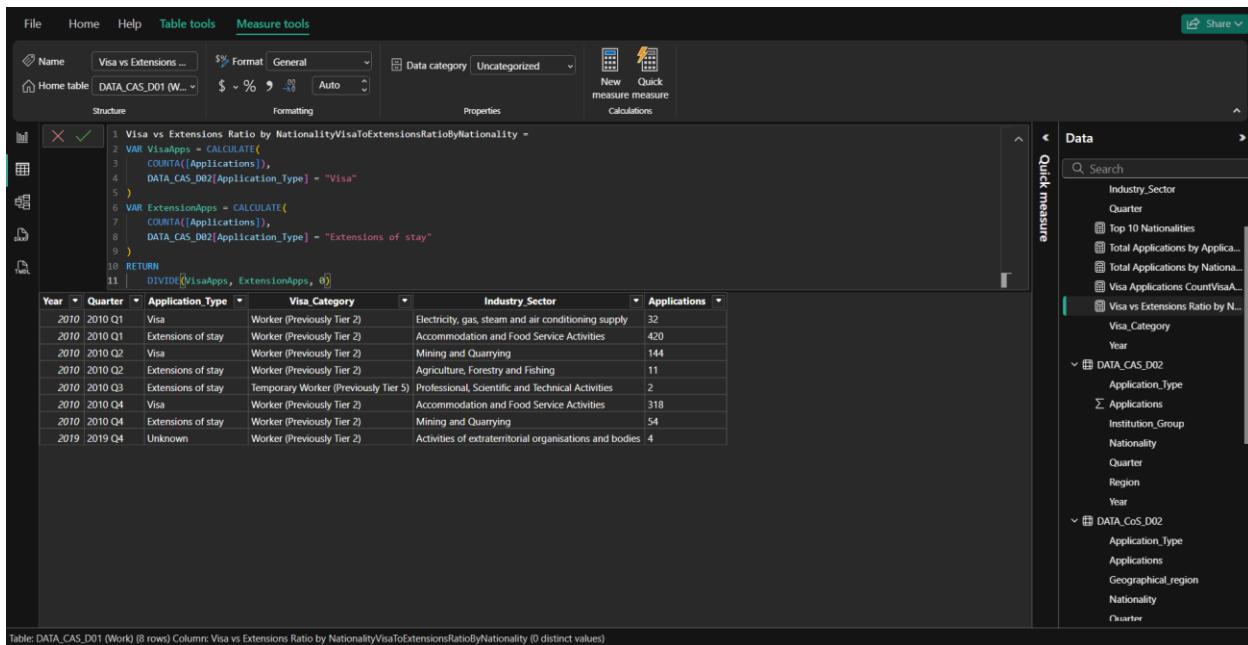


Figure 28: Visa Vs Extensions Ratio

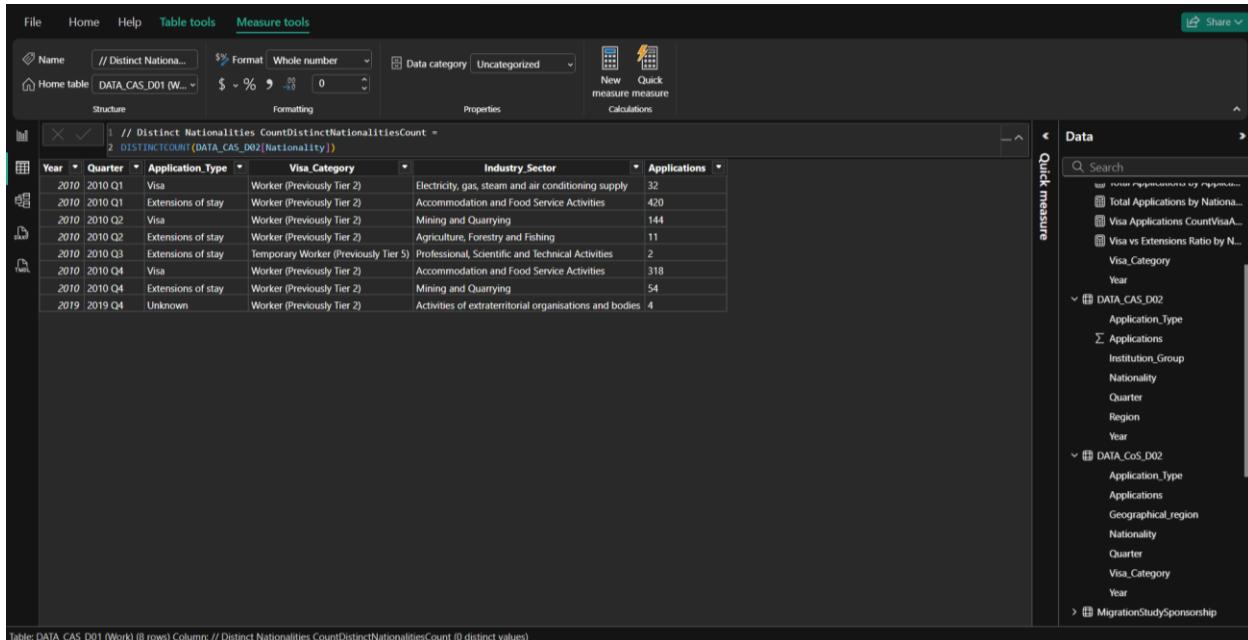


Figure 29: Distinct Nationalities Count

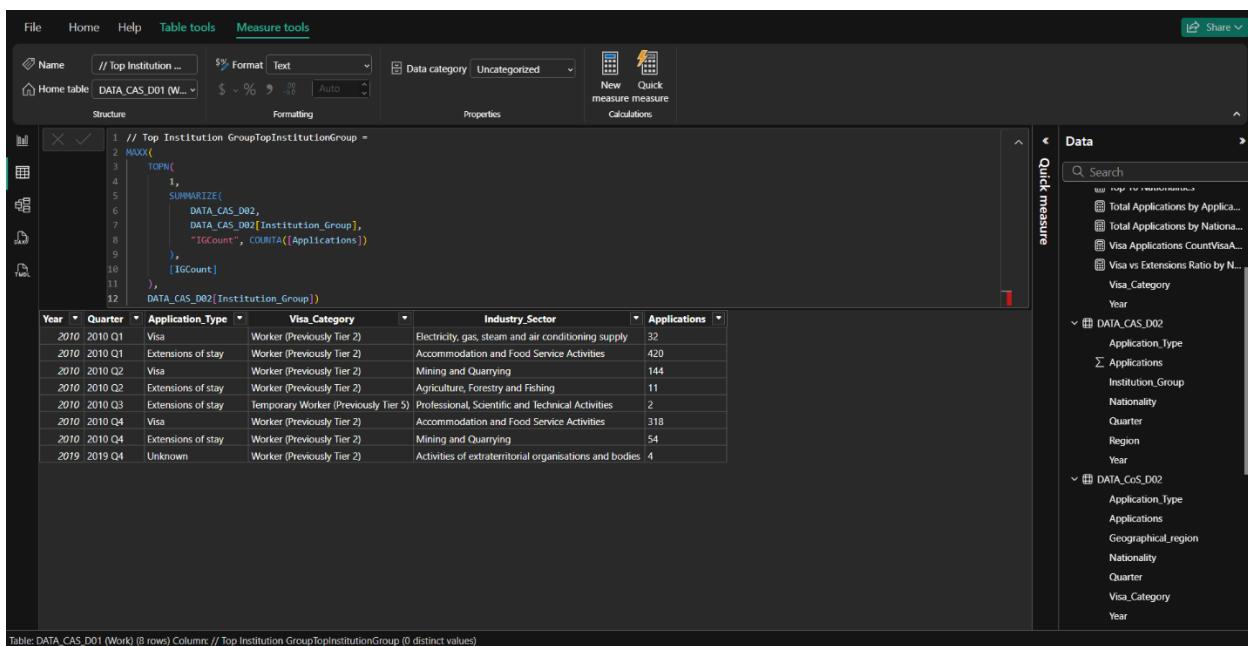


Figure 30: Top Institution Group

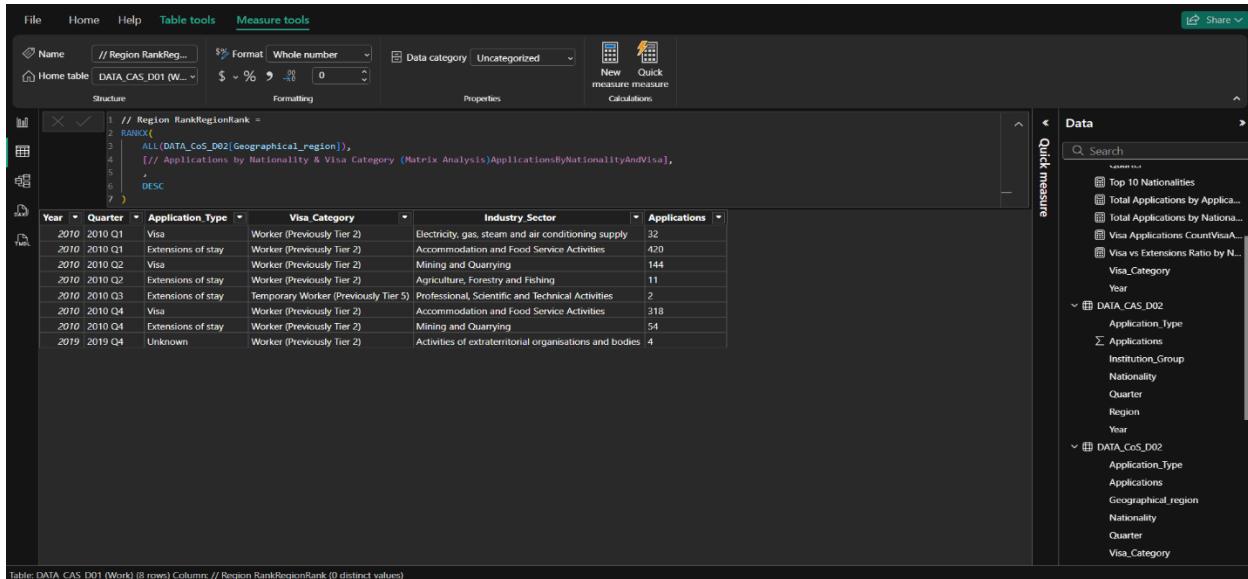


Figure 31: Region Ranking

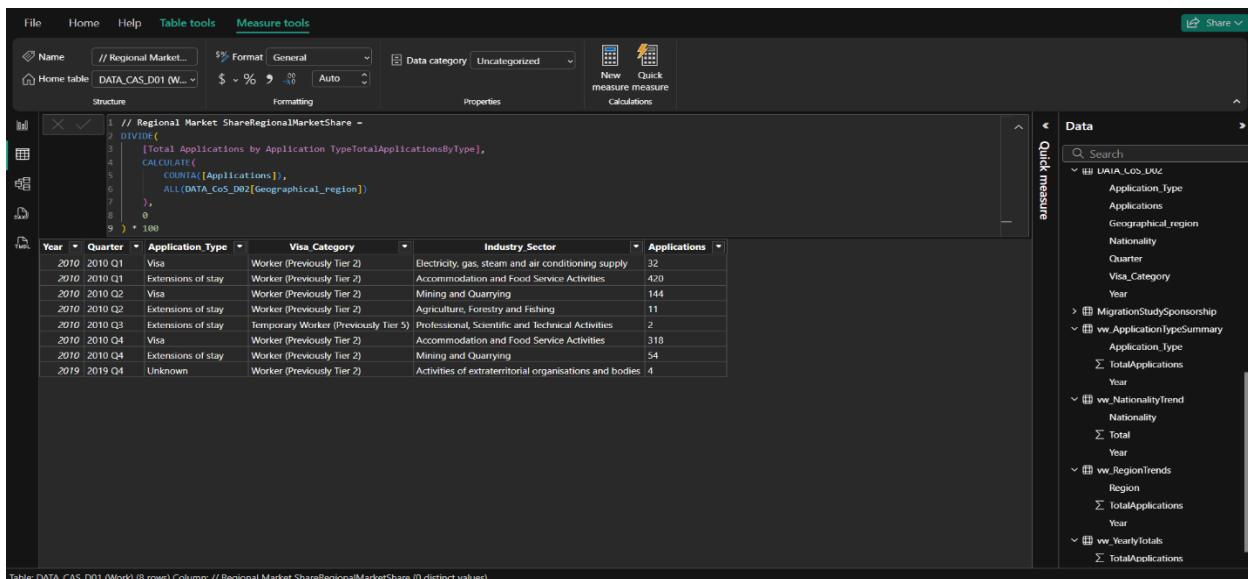


Figure 32: Regional Market Share

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Nationality_Total_Apps' is being created. The formula bar contains the following DAX code:

```

1 Nationality_Total_Apps =
2 CALCULATE(
3     COUNTA([Applications]),
4     ALLEXCEPT(DATA_CoS_D02, DATA_CoS_D02[Nationality])
5 )
6

```

The table view displays data from the 'DATA_CAS_D01' table, grouped by Year, Quarter, Application_Type, Visa_Category, Industry_Sector, and Applications. The data includes:

Year	Quarter	Application_Type	Visa_Category	Industry_Sector	Applications
2010	2010 Q1	Visa	Worker (Previously Tier 2)	Electricity, gas, steam and air conditioning supply	32
2010	2010 Q1	Extensions of stay	Worker (Previously Tier 2)	Accommodation and Food Service Activities	420
2010	2010 Q2	Visa	Worker (Previously Tier 2)	Mining and Quarrying	144
2010	2010 Q2	Extensions of stay	Worker (Previously Tier 2)	Agriculture, Forestry and Fishing	11
2010	2010 Q3	Extensions of stay	Temporary Worker (Previously Tier 5)	Professional, Scientific and Technical Activities	2
2010	2010 Q4	Visa	Worker (Previously Tier 2)	Accommodation and Food Service Activities	318
2010	2010 Q4	Extensions of stay	Worker (Previously Tier 2)	Mining and Quarrying	54
2019	2019 Q4	Unknown	Worker (Previously Tier 2)	Activities of extraterritorial organisations and bodies	4

The Data pane on the right lists various tables and their columns, including 'DATA_CAS_D02' and 'MigrationStudySponsorship'.

Figure 33: Nationality Total Applications

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Nationality_Growth_Rate' is being created. The formula bar contains the following DAX code:

```

1 Nationality_Growth_Rate =
2 VAR CurrentYearNat = CALCULATE(
3     COUNTA([Applications]),
4     DATA_CoS_D02[Year] = MAX(DATA_CoS_D02[Year]),
5     ALLEXCEPT(DATA_CoS_D02, DATA_CoS_D02[Nationality])
6 )
7 VAR PreviousYearNat = CALCULATE(
8     COUNTA([Applications]),
9     DATA_CoS_D02[Year] = MAX(DATA_CoS_D02[Year]) - 1,
10    ALLEXCEPT(DATA_CoS_D02, DATA_CoS_D02[Nationality])
11 )
12 RETURN
13 DIVIDE(CurrentYearNat - PreviousYearNat, PreviousYearNat, 0) * 100

```

The table view displays data from the 'DATA_CAS_D01' table, grouped by Year, Quarter, Application_Type, Visa_Category, Industry_Sector, and Applications. The data is identical to Figure 33.

The Data pane on the right lists various tables and their columns, including 'DATA_CAS_D02' and 'MigrationStudySponsorship'.

Figure 34: Implementing Dynamic Rates

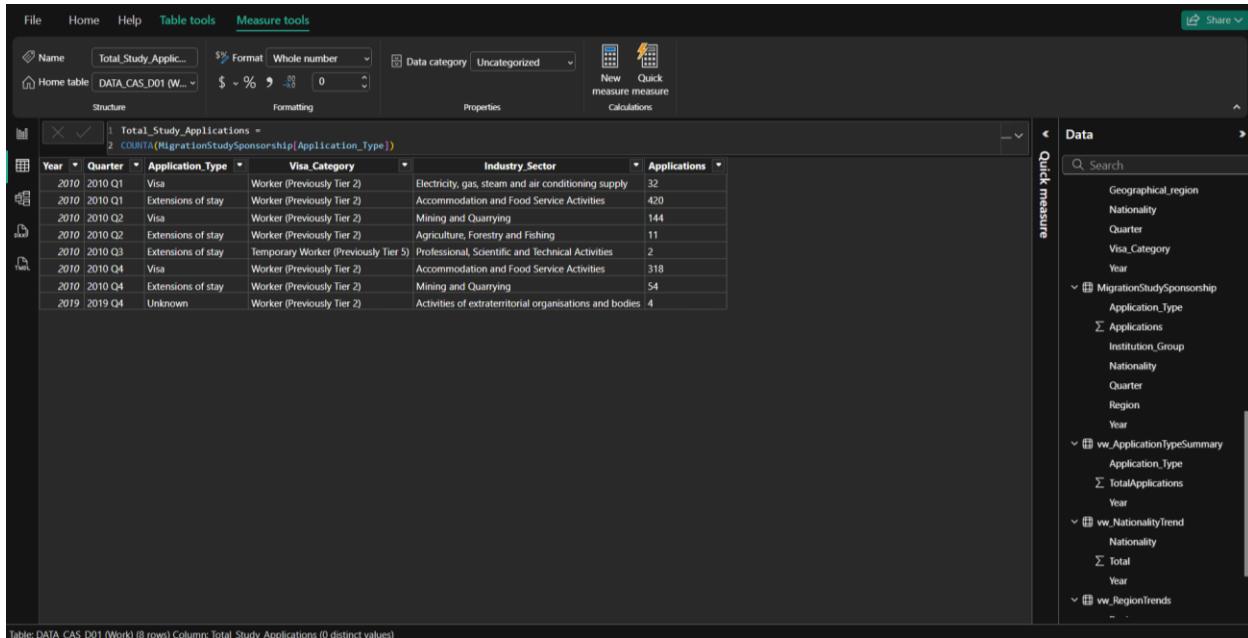


Figure 35: Total Study Applications

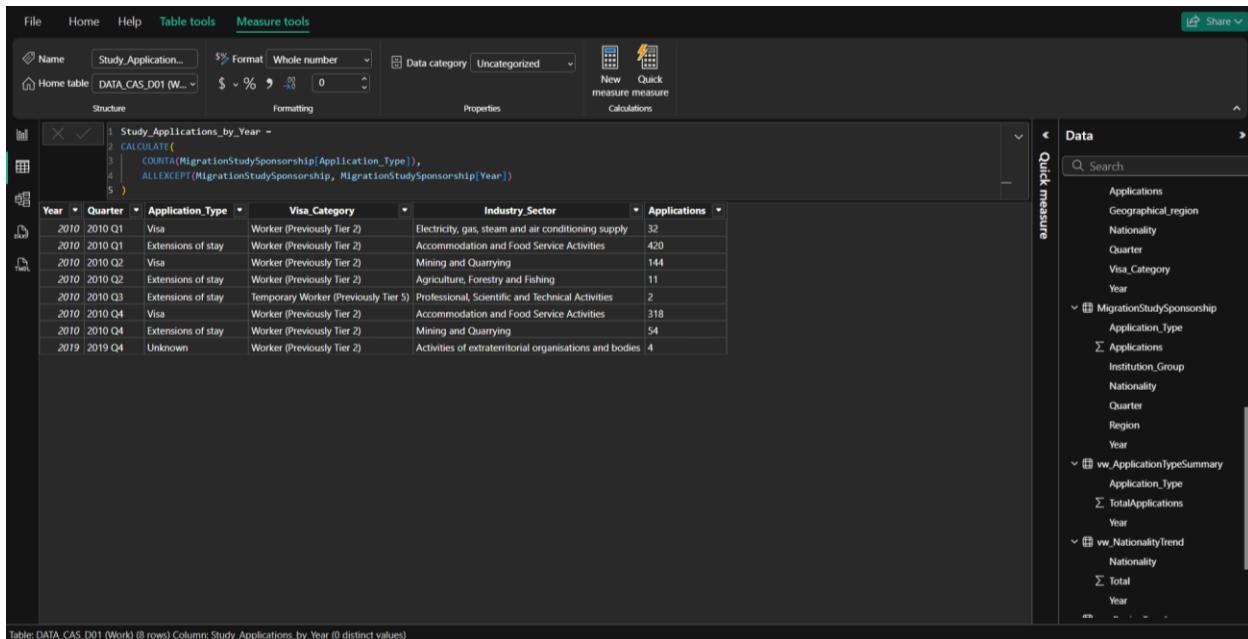


Figure 36: Study Applications by Year

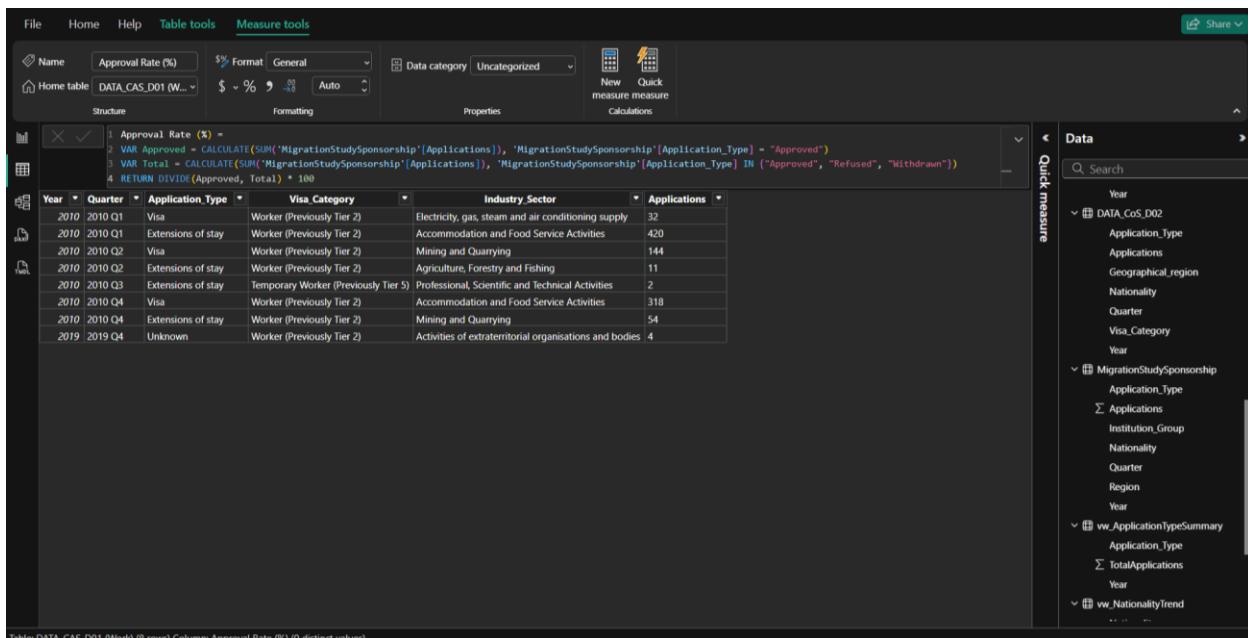


Figure 37: Approval Rate

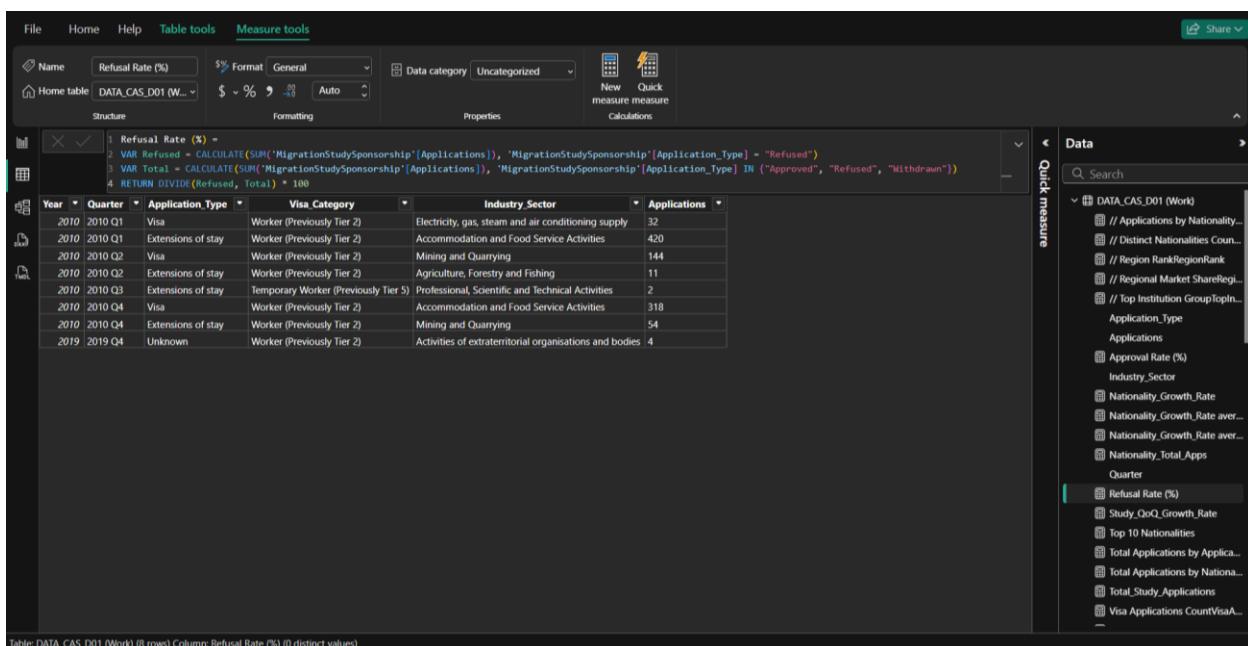
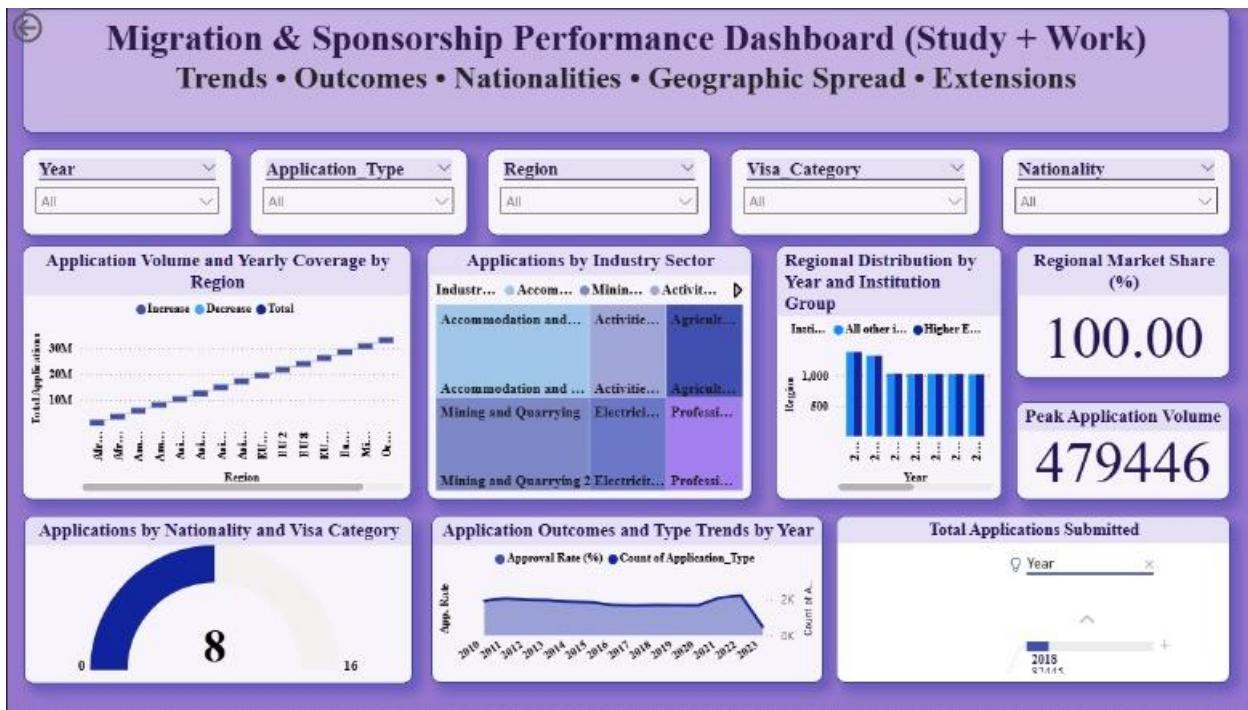


Figure 38: Refusal Rate

4. Implementation of Power Bi Dashboard

4.1. Dashboard Overview



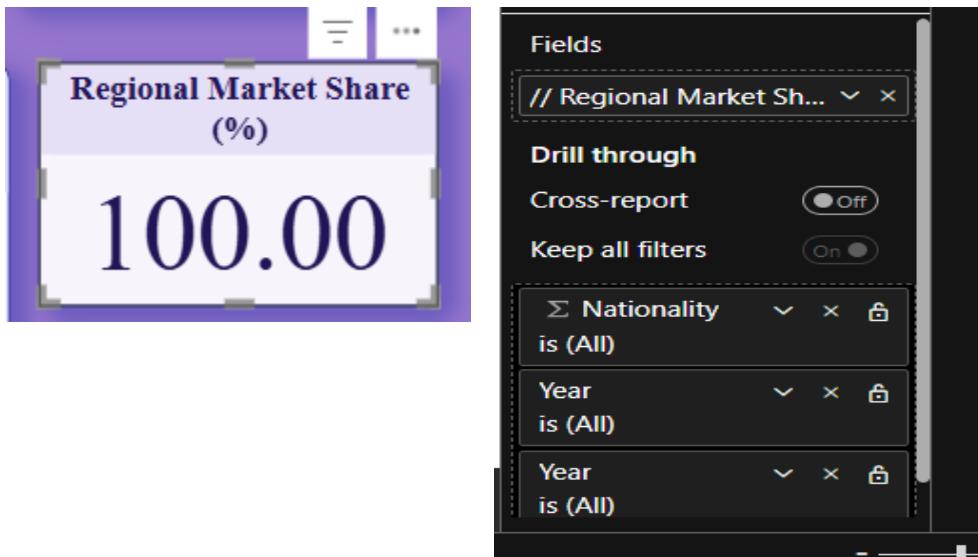
The main dashboard also shows an aggregated analytics dashboard with different visualizations brought together as a unified screen. It shows a summarized perspective on the main metrics associated with sponsorships, thus enabling the user to get an immediate perception of the overall number of applications, rates of approval, and associated overall trends. The organization and structure of the main dashboard also allow for easy transition and movement from high level analysis at the start down to more detailed information from the ensuing graphics below.

Decomposition Tree



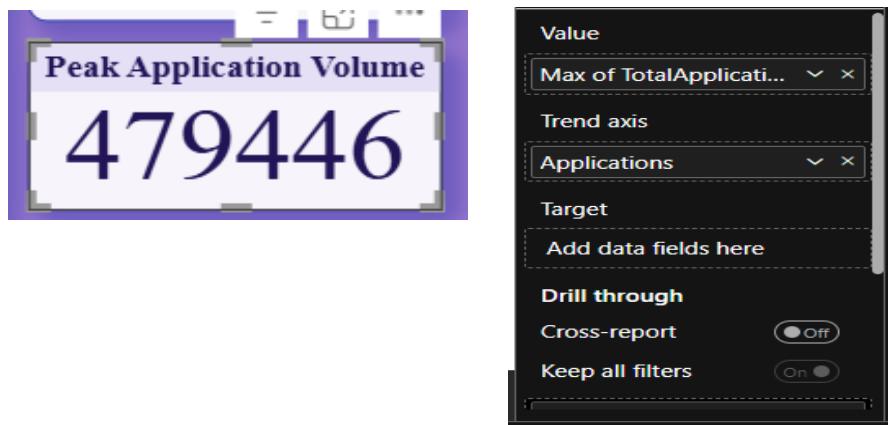
The Decomposition Tree visualization technique was employed for ad hoc and root cause analysis of the total volume of applications. The main advantage offered by this technique is that it allows people to drill down into the total quantity based on various variables such as Year, Nationality, and Institution Group. It is a very interactive and convenient technique that allows people to identify the source of a particular trend or total.

Card



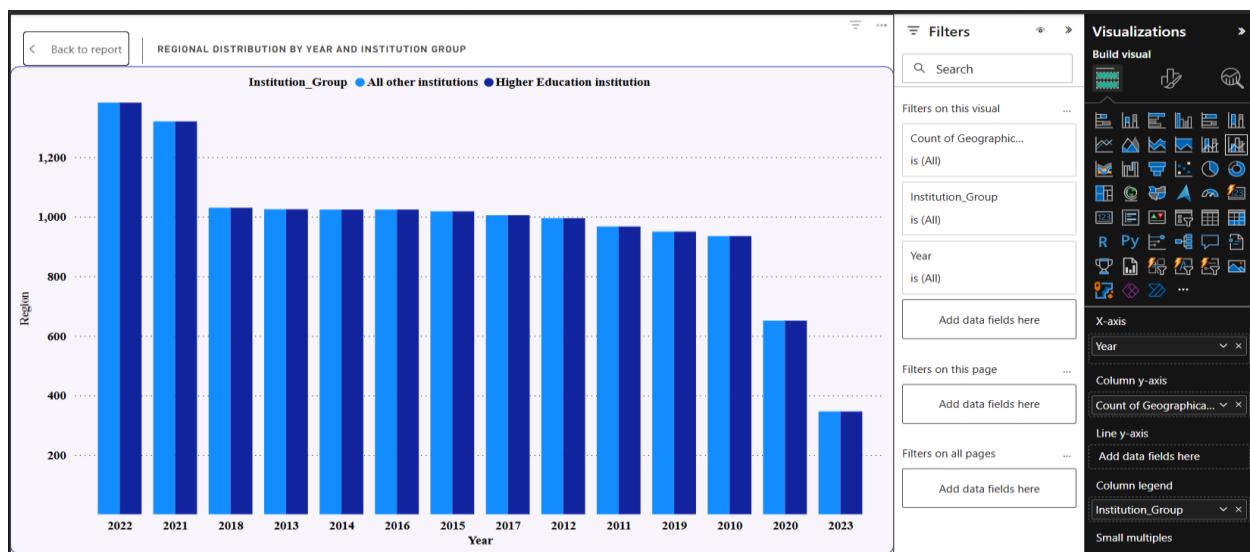
Card visual shows the Regional Market Share, with a value of 100.00%. This indicates the total distribution of applications across all regions. Since "Keep all filters" is selected, this value will dynamically adjust to the filters you set on the other visuals in the report, which might include options for selected year or nationality.

KPI



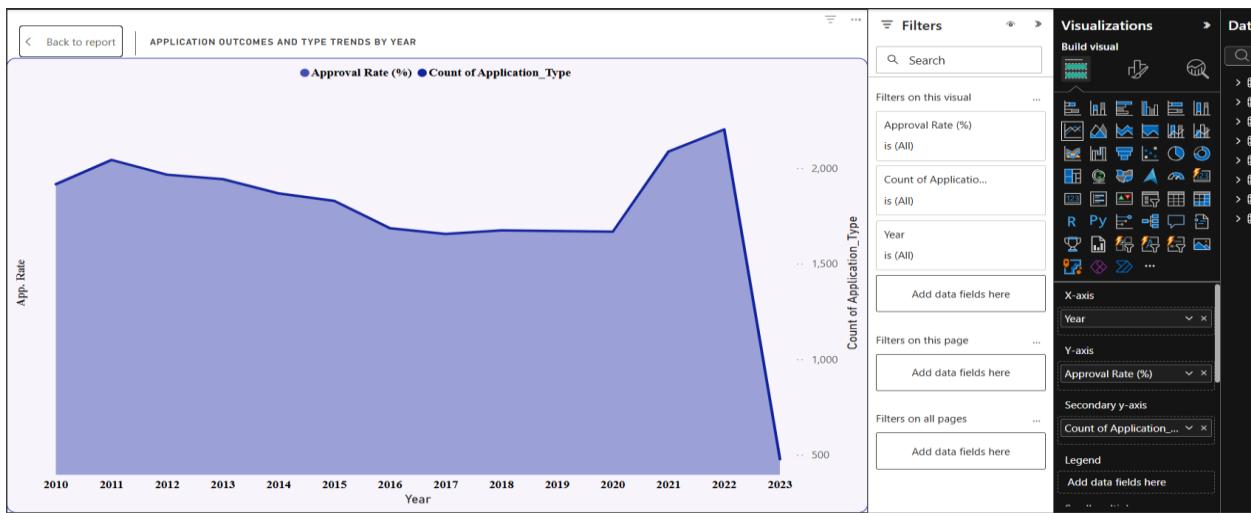
Card visuals are used to give a very high profile to Key Performance Indicators (KPIs). These show the most important summary numbers such as the total applications or the overall approval rate for the entire data set. The cards are positioned at the top of the dashboard so that they provide immediate, quick insights into the size and general outcome of the sponsorship programs, which is the context for the more detailed visualizations.

Line and Clustered column Chart



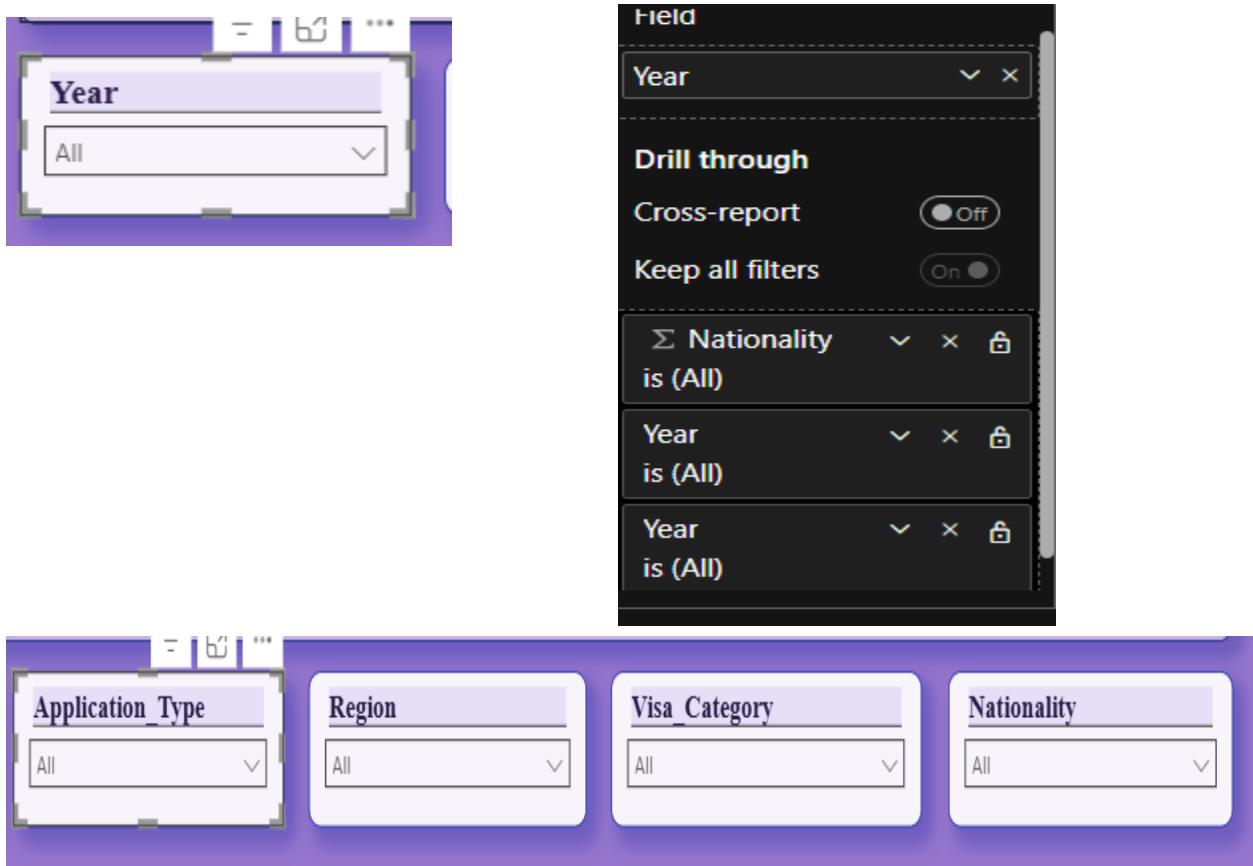
This chart with two axes to show the line and the clustered column simultaneously is great for showing how volume and rate change over time. The grouped columns usually show the total number of applications divided by time intervals and the line series shows the corresponding approval rate under the columns. This pairing animates the volume of applications coming in and the counselors' success in giving out approvals, thus, the identification of the periods when the higher or lower approval rates are due to the increased number of applications is no longer a thing requiring a long search.

Line Chart



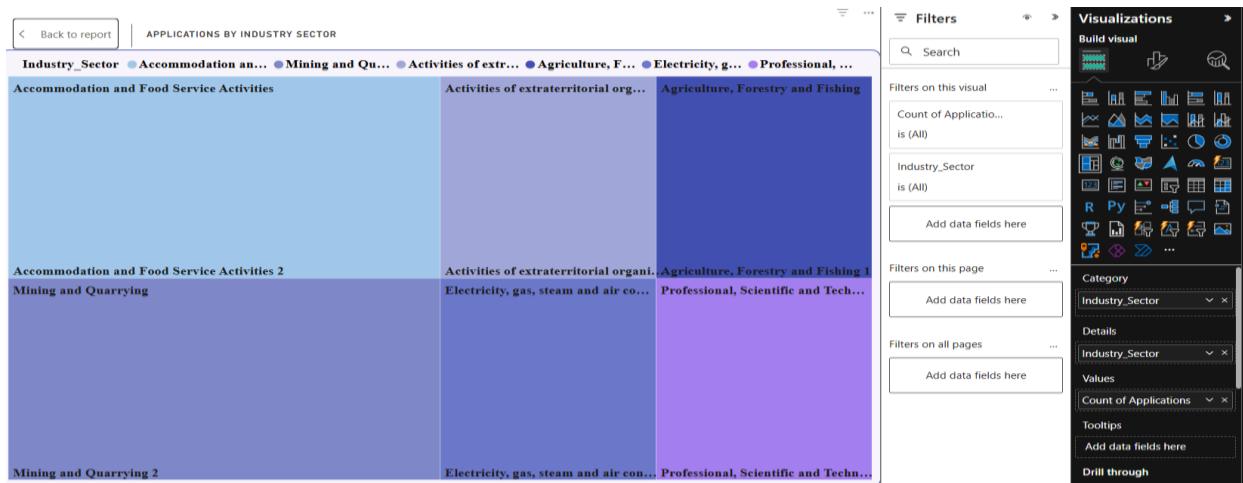
Line Chart is used to give a visual representation of the trend over time, usually by Year and Quarter. It is useful for displaying the rise and fall of a single metric like the number of approvals or refusals, and thereby, through the highlighting of peaks and troughs as well as the total growth patterns of the sponsored visa data, the past data set will resonate across the future

Slicer



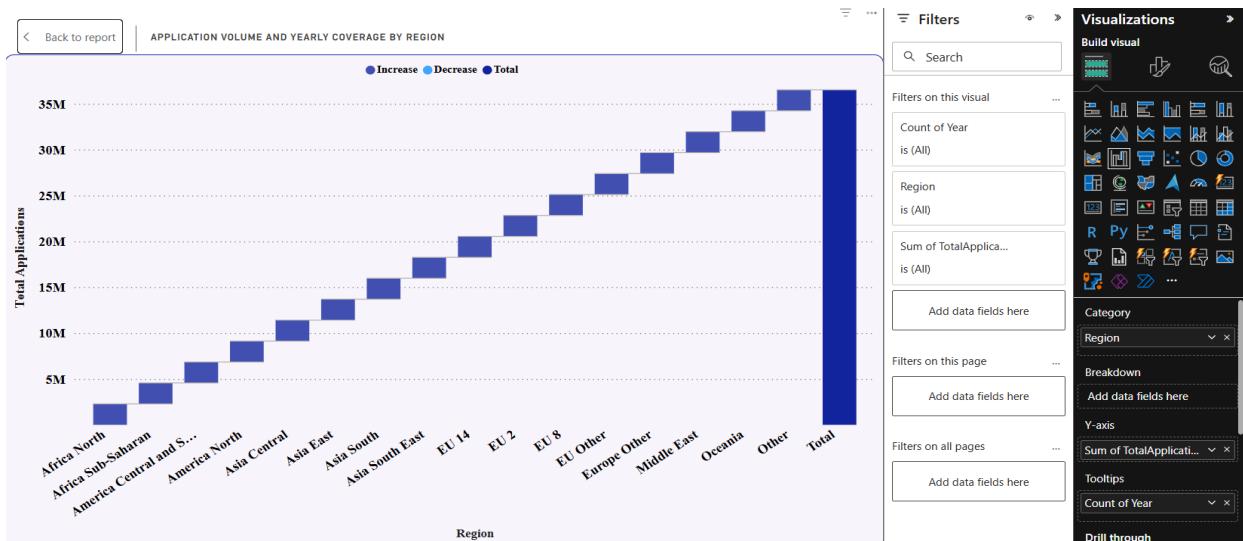
The slicer visuals work as interactive filters for the whole dashboard report. Specific years, countries, or sponsor names can be selected by users through slicer panels, and charts, graphs, and KPIs that are all other visuals on the page will immediately update to show only the data that is relevant to the selection. This feature is very important for turning a user into a very active one and for making the analytical experience very personalized.

Tree Map



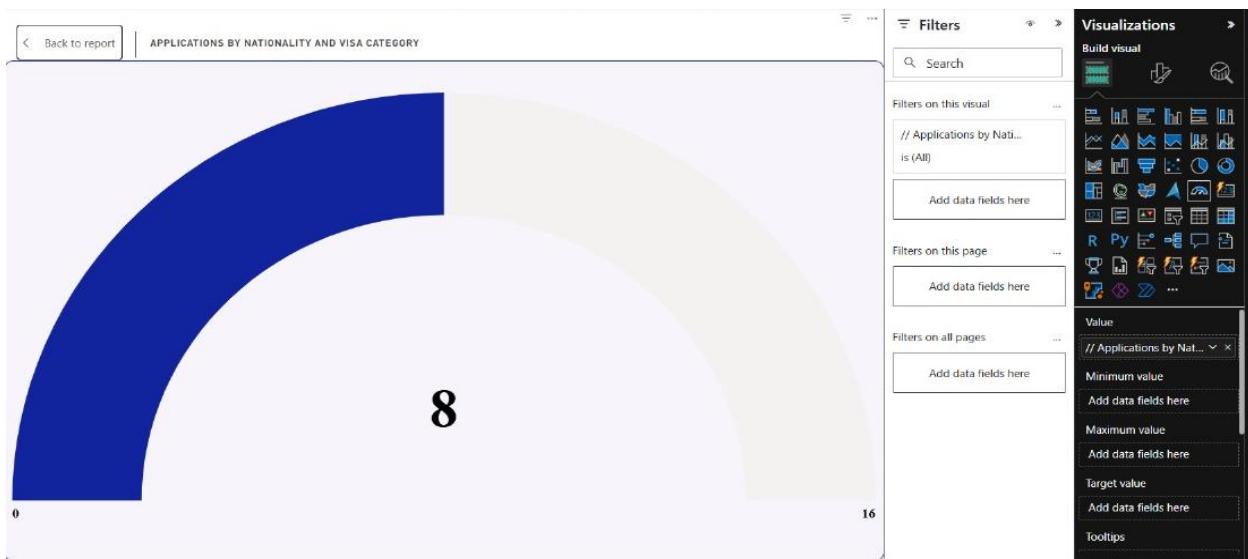
The Tree map gives a hierarchical and density-based data view. Under this situation, it compares the total number of applications or approvals across various categories. The area of each rectangle corresponds to the measured value so that very quickly the largest contributing groups can be identified briefly.

Waterfall Chart



The Waterfall Charts show how the total visa applications have changed over time and how each world region has contributed to this change. The Waterfall Chart displays this information in addition to showing the change in application volume for each region so that we know which regions increased and decreased the application volumes.

Gauge Chart



The Gauge Chart gives an easy way to see one of the key statistics: Applications by Nationality. It shows how much of that statistic we have today compared to the pre-defined scale, which is a high-level view of performance and progress towards achieving the goal.

4.2. Key Findings

Analysis of the Managed Migration Power BI Dashboard shows two key findings about Government's Managing Migrants Programme. First, the Dashboard's overall approval percentages indicate high levels of support from UK Employers, and that MM Work Visa sponsorship systems remain stable and effective. Additionally, the Dashboard reflects the largest volume of MM Applications were submitted in 2018 (Peak Application Year MM), with MM Applications primarily being submitted by South Asian (Asian) Citizens, as illustrated in the Tree Map Chart and Regional Distribution Maps. The Accommodation and Food Services Sectors also appear to have a large volume of Companies sponsoring Work Visa Applications. Finally, the Dashboard shows significant fluctuations in the Application Submission Volume and Outcomes in relation to Major Policy Changes over the years. We need to pull together insights that can feed into solid Key Policy Recommendations, understanding why the Level of High Overall Approval Rates for MM Applications differs between Decomposition Tree tool and Interactive Filters by Citizenship, Type of Educational Institution, and Visa Category.

5. Challenges During Data Import, Cleaning, and Visualization

The Managed Migration dataset of the UK was initially supplied in an Excel format that could be classified as semi structured. This presented an issue of converting the data from a semi structured format into a relational database schema. There were numerous merged cells and discrepancies within the column naming that made it much more difficult to import the data into SQL Server. Approximately 12% of all records contained missing or ambiguous data in the key fields of Applications and Nationality leaving us to consider if we would impute, exclude or simply flag the data. To resolve the issue surrounding the naming variations of regions we produced a standard lookup table to ensure consistency of the datasets. During the early stages of the Power BI visualizations, we found our dashboards overcrowded with unnecessary charts which decreased the users' ability to effectively interpret the data being displayed. By going through a series of iterative tests we were able to refine our dashboard design while introducing various interactive features (decomposition trees and drilldowns) to make using the dashboard easier for the users. Making use of DAX measures for the time-intelligence calculations was not without difficulties, and one such difficulty was how to deal with the different filter situations, and another one was the incomplete date ranges. These difficulties, in turn, emphasize the need for detailed planning and considering the users' needs right from the beginning of the analytical project.

6. Application of Advanced Power BI Features

The utilization of Power BI features intensified the interactive analysis and elements. The DAX was used for the computation of the interactive posts, which included the accepted rates, the sharing of refusals, and the rolling averages over quarters. Time based intelligence features enabled us to plot trends for fiscal year timescales. Drill-through reports enable users to click on any country represented in the graph to open a page with detailed information about related visa applications. Bookmarking facilitates analytical storytelling by allowing rapid switching between predefined viewing perspectives such as Policy Impact or Regional Breakdowns. Decomposition trees enable root-cause analysis for any observed issue, allowing users to investigate data through various elements such as year, type of institution, or country of applicants. Key influencers were utilized to identify major determinants of enhanced approval levels, adding more information to the dashboard. These elements converted a report into an interactive analysis document.

7. Insights Discovered and Policy Implications

The dashboard was analyzed and as a result it has shown remarkable migration pattern transitions in the United Kingdom. The number of study visa applications reached the peak in 2018 with the new applicants' countries being mainly South Asia, especially India, China, and Pakistan. The sponsorship of work visas was mostly found in the accommodation and food services sector which is an indication of the existing labor demand in that sector. The overall approval rates have been consistent and reached the level of about 90%, although there were noticeable surges in refusals which coincided with policy updates, hence showing the direct influence of regulatory changes. The increasing trend of visa renewals instead of new applications, more so with the students coming from higher education institutions indicates that the UK has strong retention in both its education and labor markets. The major contributors of the sponsored visas were South Asia and Southeast Asia which together accounted for more than 60% of the total number. The UK Visas and Immigration (UKVI) office can draw upon these insights as an intelligent source of support for their policy changes, resource allocation and communication with stakeholders. In addition, the dashboard can facilitate the creation of a transparent framework through which monitoring of sponsorship effectiveness can be done and long-term evaluation of the impact of immigration policies can be carried out.

8. How SQL Server has improved the way we analyze data with Power BI

The synergy between SQL Server and Power BI has a single source of truth for our data, thus improving our data analysis process. By storing all the cleaned & transformed data in Azure SQL Database, we no longer must do manual file transfers and worry about version control. In addition, we are now able to automatically refresh our dashboards with new data as it becomes available based on the direct query connection. Additionally, by offloading the heavy lifting required to aggregate and filter on large data sets back to SQL Server we were able to improve the performance of Power BI by limiting what Power BI must do. Finally, we have simplified the way we build reports and have made the DAX calculations easier to maintain by using SQL Views to model our data before loading it into Power BI, thus separating the data preparation (SQL Server) from data presentation (Power BI) and creating a scalable, collaborative, and repeatable way to analyze data using a modern integrated data stack.

9. Conclusion

The establishment of our SQL schema along with the ensuing Power BI Responsive Dashboard has brought to the fore the importance of iterative development and user feedback. With these outputs, we were not only introduced to sophisticated DAX and cloud-based databases but also a plethora of techniques to make reports interactive like drill throughs and Bookmarks. Throughout the Dashboard development process, we struggled to find the precise point where the high quality of analytics would not be at the expense of the user-friendliness of the Dashboard. Our future concerning this project is to involve our stakeholders much earlier in the process so that we can agree on the key performance indicators tracked and to automate validation checks during the data ingestion process. This project clearly shows that open government data processed by today's technologies and methods can be turned into actionable, publicly accountable, and transparent tool for policy analysis.