








# Product Licenses Sales Analytics

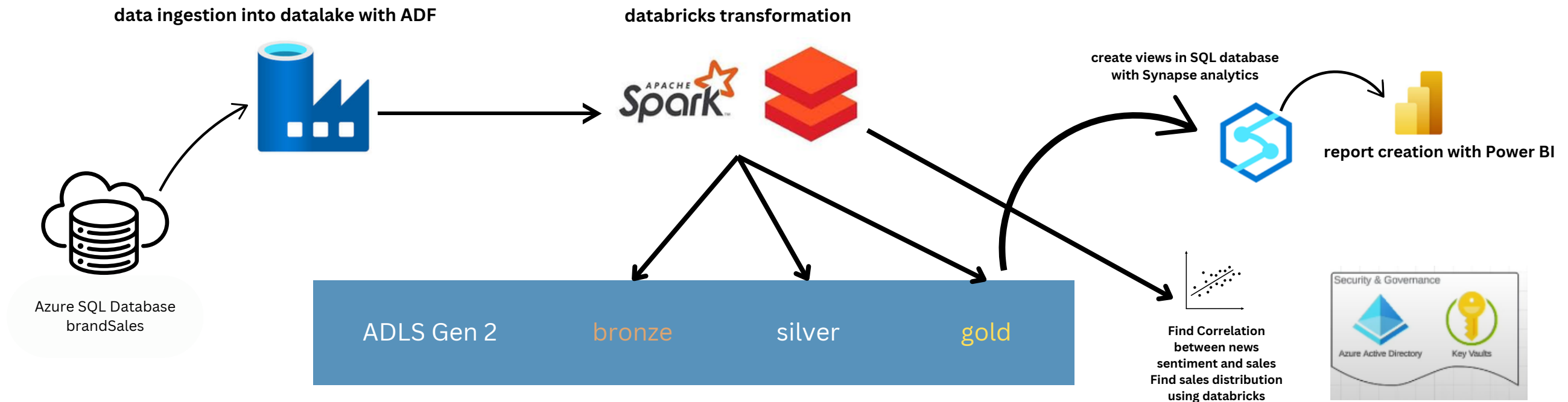
using Microsoft Azure Platform



## 1. New Resource Group Created with following resources -

Name ↑	Type
 adfproj2e	... Data factory (V2)
 brandSales (partnerandcertification/brandSales)	... SQL database
 partnerandcertification	... SQL server
 proje2e	... Storage account
 proje2e-databricks	... Azure Databricks Service
 proje2e-keyvault	... Key vault
 synapse-proje2e	... Synapse workspace

## 2. Architecture




### 3. Dataset Description

A software development company sells software licenses worldwide.  
It caters to different Geographies / Regions.  
They are being sold to companies belonging to several industries.

Each product is built under a particular brand, each brand can be considered a product line.  
The company sells via partner companies acting as resellers. Each partner company has sales employees called as partners.  
Each partner trains under a certification for every product under the brand (combination of unique certification id and certification name pair).  
On passing, they become eligible to conduct sales activities for that product licenses.

License types are standard, professional and premium.

	brand	certification_id	certification_name	lastest_version
1	Netvibes	CERT081	Leadership Foundations-Netvibes	15-08-2017
2	3Dvia	CERT094	Data Analysis Applications-3Dvia	03-12-2017
3	3Dvia	CERT055	Business Intelligence Foundations-3Dvia	30-09-2017

Data is in SQL database 'brandSales'

- Latest version of each product given by combination of certification\_id and certification\_name is maintained in table 'certificationLatestVersion'

	brand	certification_id	certification_name	lastest_version
1	Netvibes	CERT081	Leadership Foundations-Netvibes	15-08-2017
2	3Dvia	CERT094	Data Analysis Applications-3Dvia	03-12-2017
3	3Dvia	CERT055	Business Intelligence Foundations-3Dvia	30-09-2017
4	3Dvia	CERT149	Analytics Foundations-3Dvia	10-02-2017
5	Geovia	CERT066	Chemical Engineering Applications-Geovia	10-06-2017
6	Simulia	CERT104	Supply Chain Strategies Applications-Simulia	08-11-2017
7	Solidworks	CERT039	Risk Management Foundations-Solidworks	02-03-2017
8	Solidworks	CERT172	Fabrication Processes Applications-Solidworks	20-08-2017
9	Netvibes	CERT042	Prototyping Techniques Applications-Netvibes	03-06-2017
10	Geovia	CERT071	Product Development Foundations-Geovia	19-08-2017

- Pass dates for each product (combination of certification\_id and certification\_name) is maintained in table ‘partnerCertificationPassDate’.

It also follows a sales channel Industry>ISE>IPE

The partner also needs to be provided target job role they can sell to.

Partner company ids are unique for each reseller company and partner id is also unique for each partner.

Additionally we have information of location such as region, country, state the partner company is allowed to market to.

Multiple certification holder signifies whether the partner holds multiple certifications.

	Brand	Certification_ID	Certification_Name	Partner_Company	Partner_Company_ID	IPE_ID	ISE_ID	Industry	IPE	ISE	Tgt_Job_Role	Partner_First_Name	Partner_ID	Partner_Last_Name	Multiple_Cert_Holder	Region	Country	State	Pass_Date
1	Solidworks	CERT038	Smart City Applications-Solidworks	Tracking Technologies	TRAC	P00109	S00101	Aerospace & Defense	Workforce Management	Smart Manufacturing	Assembling Engineer	FirstName0	TRAC16117	LastName0	FALSE	Asia	India	Himachal Pradesh	2018-05-20T14:03:38.148Z
2	Netvibes	CERT057	Automation Foundations-Netvibes	Nexus Engineering	NEXUS	P00112	S00115	Consumer Packaged Goods Retail	Maintenance Scheduling	Data-Driven Decision Making	Program Manager	FirstName1	NEXUS03086	LastName1	FALSE	Asia	India	Himachal Pradesh	2017-06-15T15:28:54.107Z
3	Netvibes	CERT011	Electrical Design Foundations-Netvibes	Dynamic Systems	DYNAM	P00114	S00102	Architecture, Engineering & Construction	Project Management	Digital Twin Technology	Engineer	FirstName2	DYNAM62697	LastName2	FALSE	Asia	Russia	Leningrad Oblast	2018-03-18T07:33:46.375Z
4	Catia	CERT081	Leadership Foundations-Catia	Nexus Engineering	NEXUS	P00103	S00103	Business Services	Quality Control Processes	Advanced Materials Development	Facility Supervisor	FirstName3	NEXUS51276	LastName3	FALSE	Asia	India	Himachal Pradesh	2018-08-13T07:29:26.822Z

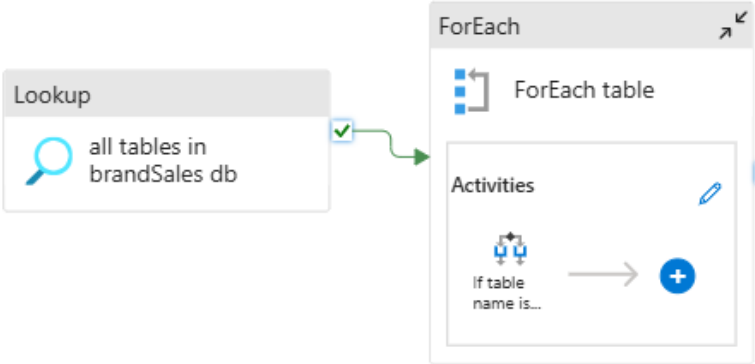
- We have a table ‘partnerlicensesales’ that stores the sales figures for years 2019, 2020, 2021, 2022 month-wise.

	Partner_ID	Certification_ID	Industry	Certification_Name	ISE_ID	IPE_ID	Partner_Company_ID	UK_ID	Jan Standard	Jan Professional	Jan Premium	Feb Standard	Feb Professional	Feb Premium	Mar Standard	Mar Professional	Mar Premium
1	TRAC16117	CERT038	Aerospace & Defense	Smart City Applications-Solidworks	S00101	P00109	TRAC	UK23874870	5	15	7	13	11	2	20	31	14
2	NEXUS03086	CERT057	Consumer Packaged Goods Retail	Automation Foundations-Netvibes	S00115	P00112	NEXUS	UK31322254	24	11	11	6	1	11	24	7	4
3	DYNAM62697	CERT011	Architecture, Engineering & Construction	Electrical Design Foundations-Netvibes	S00102	P00114	DYNAM	UK28280210	19	8	7	11	5	12	24	23	14

- We have a table ‘newssentiment’ that stores the news sentiment data (positive, neutral, negative) year and month wise for each industry, each egion, each country and each state for years 2019, 2020, 2021, 2022.

	Region	Country	State	Industry	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	year_name
1	Africa	Egypt	Alexandria	Aerospace & Defense	negative	positive	positive	neutral	positive	negative	positive	negative	positive	neutral	negative	neutral	2021
2	Africa	Egypt	Alexandria	Cities & Public Services	negative	neutral	positive	positive	neutral	positive	positive	positive	neutral	negative	neutral	neutral	2021
3	Africa	Egypt	Alexandria	Consumer Packaged Goods Retail	positive	neutral	positive	positive	neutral	neutral	negative	positive	positive	neutral	negative	negative	2021

3. Copy from SQL DB to ADLS container bronze-partnersales in parquet format (SCD type 1)



copy-dbRaw2BronzeAdls > ForEach table > If table name is highwatermarkdate > False activities

General Source Sink Mapping Settings User properties

Source dataset \* brandSales\_raw Open + New Preview data Learn more

Dataset properties

Name	Value
schema	@item().SchemaName
table	@item().TableName

Use query ☒ Table ☐ Query ☐ Stored procedure

Query timeout (minutes) 20

Isolation level Select...

Partition option ☒ None ☐ Physical partitions of table ☐ Dynamic range

Please preview data to validate the partition settings.

Additional columns

Name	Value
last_copied	@utcnow()

Location: bronze-partnersales / sales

Search blobs by prefix (case-sensitive)

Name

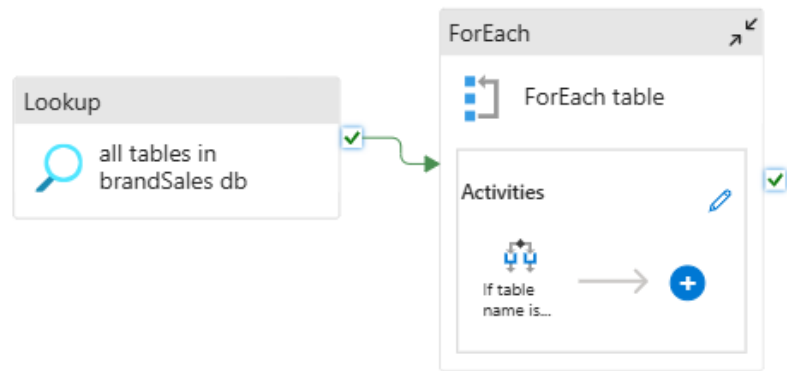
- ☐ [-]
- ☐ certificationLatestVersion
- ☐ newssentiment
- ☐ partnerCertificationPassDate
- ☐ partnerlicensesales



- ☐ bronze-partnersales
- ☐ gold-partnersales
- ☐ silver-partnersales
- ☐ synapse-workspace



3. Copy from SQL DB to ADLS container bronze-partnersales in parquet format (SCD type 1)



Location: [bronze-partnersales](#) / sales

Search blobs by prefix (case-sensitive)

Name
<input type="checkbox"/> [..]
<input type="checkbox"/> certificationLatestVersion
<input type="checkbox"/> newssentiment
<input type="checkbox"/> partnerCertificationPassDate
<input type="checkbox"/> partnerlicensesales

- ☐ bronze-partnersales
- ☐ gold-partnersales
- ☐ silver-partnersales
- ☐ synapse-workspace

copy-dbRaw2BronzeAdls > ForEach table > If table name is highwatermarkdate > False activities

Copy data

Copy table to parquet in adls

General Source Sink Mapping Settings User properties

Source dataset \* brandSales\_raw

Dataset properties

Name	Value
schema	@item().SchemaName
table	@item().TableName

Use query ☒ Table ☐ Query ☐ Stored procedure

Query timeout (minutes) 20

Isolation level Select...

Partition option ☒ None ☐ Physical partitions of table ☐ Dynamic range

Please preview data to validate the partition settings.

Additional columns

Name	Value
last_copied	@utcnow()



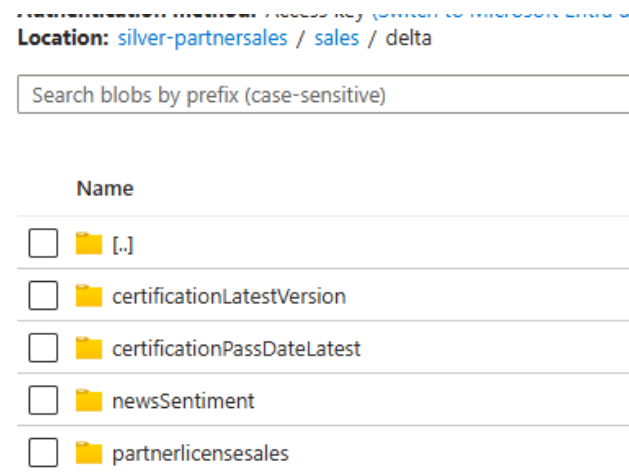
4. In Azure Databricks, we have 5 notebooks for transformation from bronze to silver to gold.

In the First 4 notebooks, we transform the 4 ADLS parquet files data and save to silver-bronzepartnersales container in delta table format

[https://github.com/brighthelianthus/brandSalesADEproject/tree/adf\\_publish/notebooks](https://github.com/brighthelianthus/brandSalesADEproject/tree/adf_publish/notebooks)

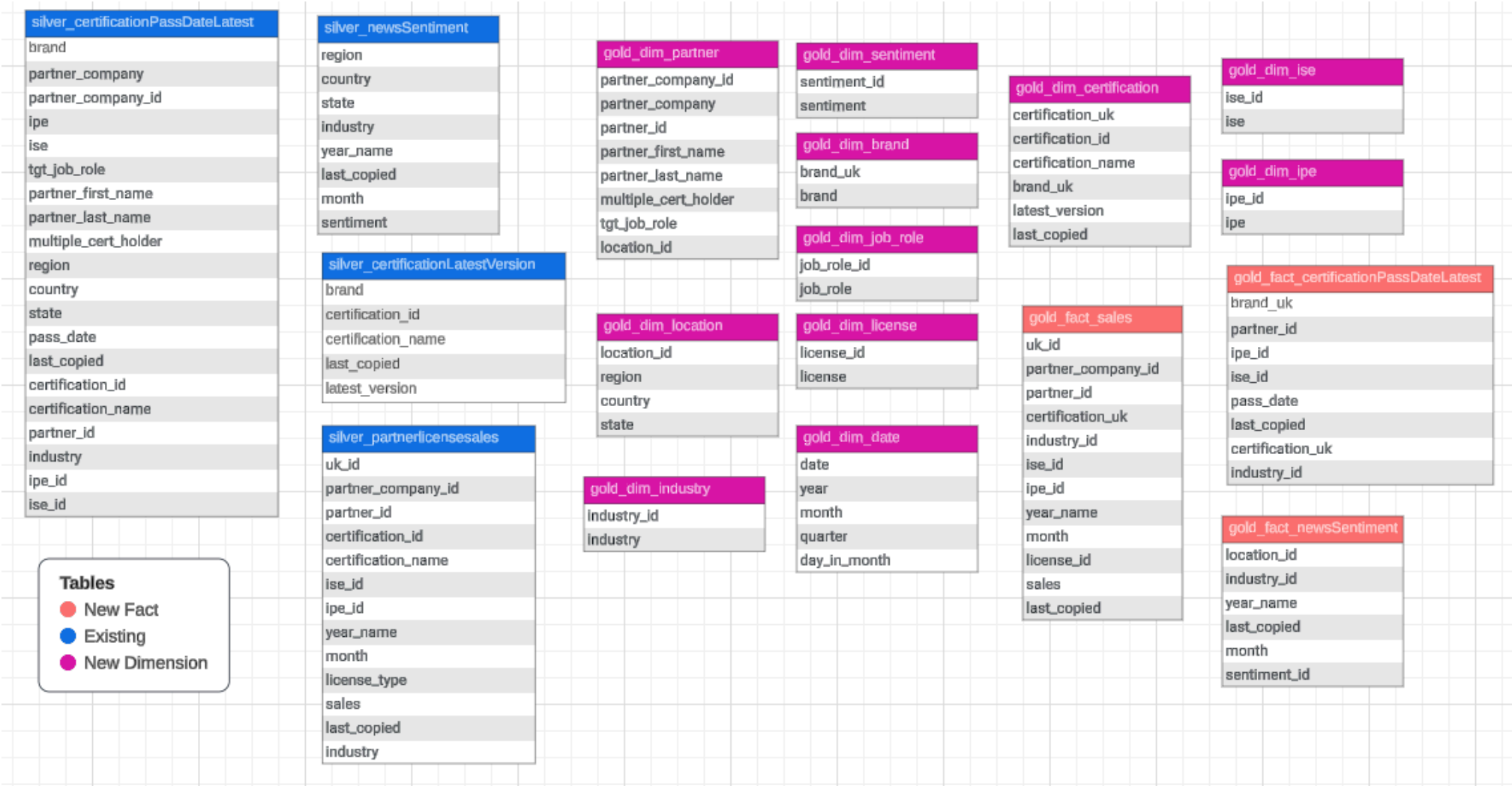


In silver-partnersales container, we are incrementally saving delta tables: certificationLatestVersion, certificationPassDateLatest, partnerlicensesales, newsSentiment



5. In the notebook silver\_to\_gold, we are creating dimension and fact tables.

[https://github.com/brighthelianthus/brandSalesADEproject/tree/adf\\_publish/notebooks](https://github.com/brighthelianthus/brandSalesADEproject/tree/adf_publish/notebooks)



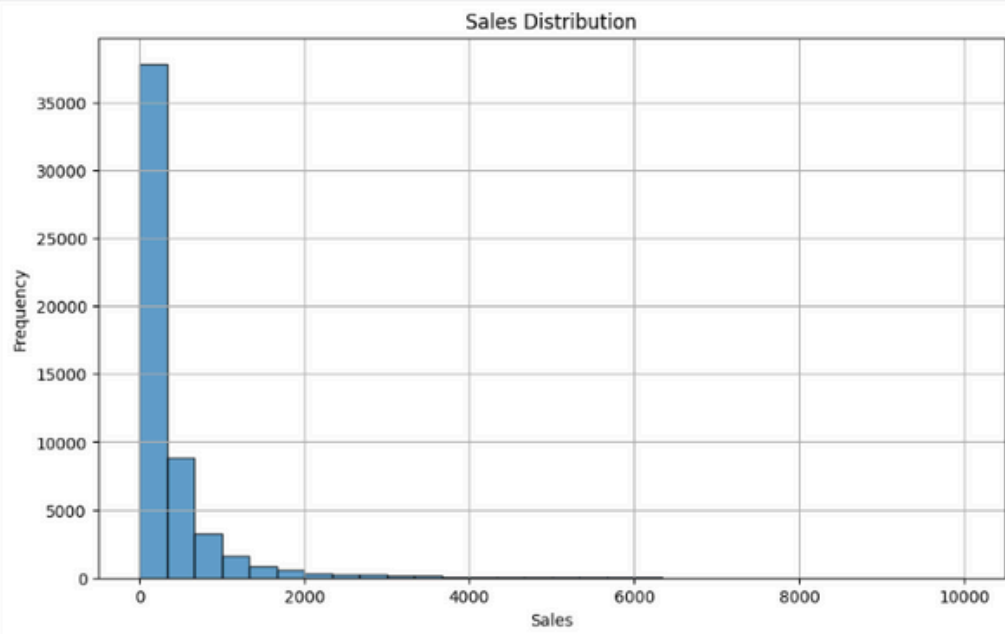
6. In the notebook `correlation_sentiment_sales.py` , we try to find if there is any correlation between the sales figures and news sentiment.

The sample dataset isn't really genuine data and hence not accurate. Still if we want to check for correlation between sentiment and sales values ...

With real data, we would expect to see some positive correlation between sentiment and sales or else weak correlation.

But with this dataset, correlation is -0.007 which signifies almost no correlation.

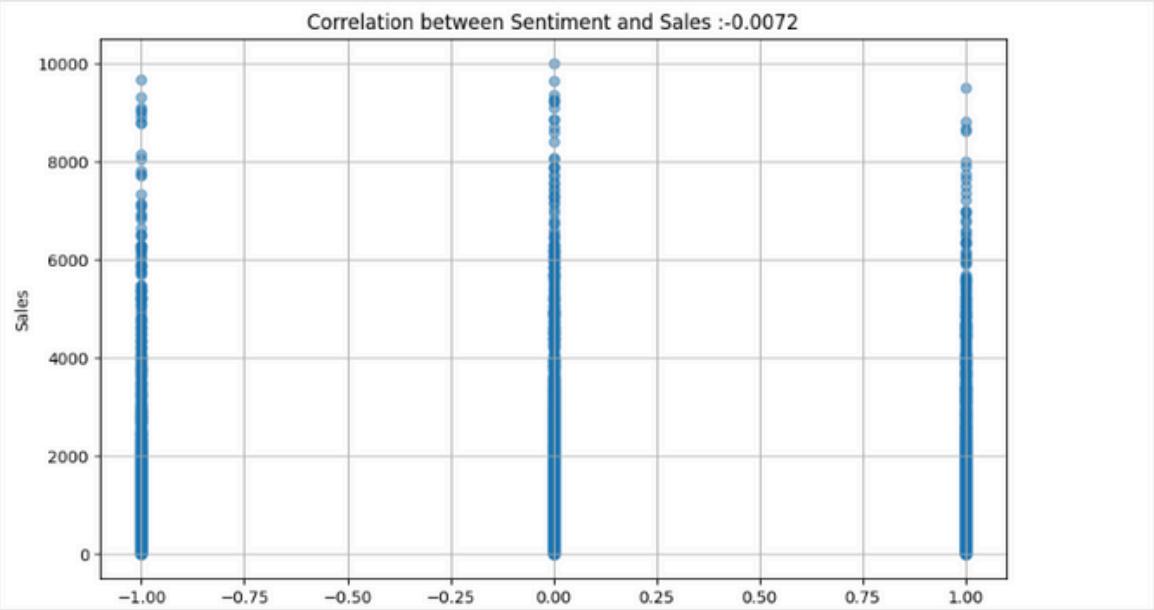
# licensesales-news sentiment correlation



	summary	sales
1	count	54480
2	mean	392.892309104258...
3	stddev	681.6815785542663
4	min	0
5	25%	84
6	50%	178
7	75%	411
8	max	9998

	sales_classification	count
1	Moderate	20729
2	High	4534
3	Low	29217

	sentiment	count
1	Negative sentiment	12168
2	Positive Sentiment	18253
3	Neutral Sentiment	24059

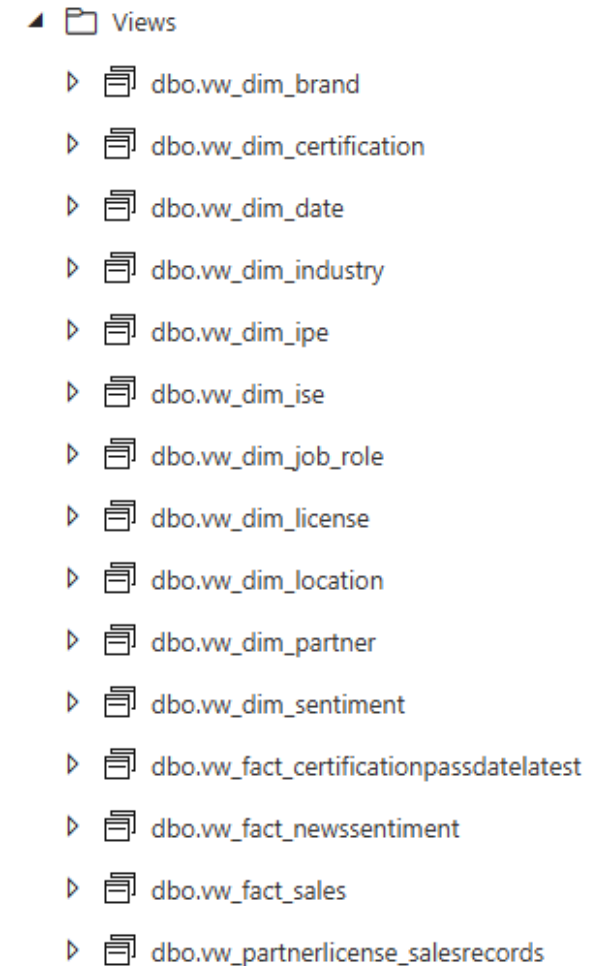


7. In Azure Synapse , linked service is created to the ADLS gold-partnersales container. We create an SQL database.

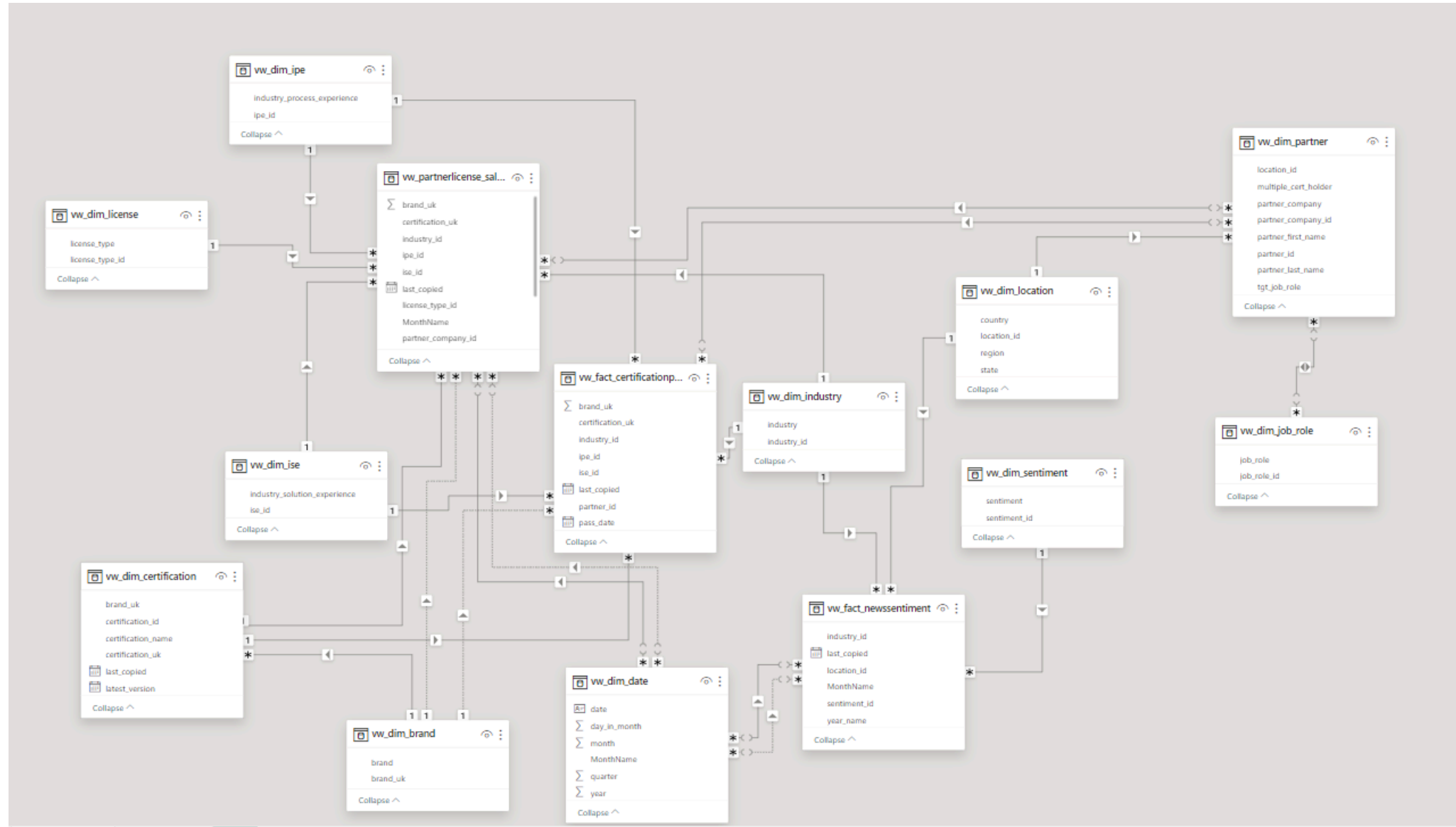
And we use it to access the gold layer delta tables and store in views. I could have created additional external table instead of directly using the gold delta format tables but here I used the gold delta tables directly and created views.

We create views of dimension tables and fact tables as required to be used in Power BI data model.

We create user for the database and give read access to it so that it can access views as data source from Power BI desktop app.

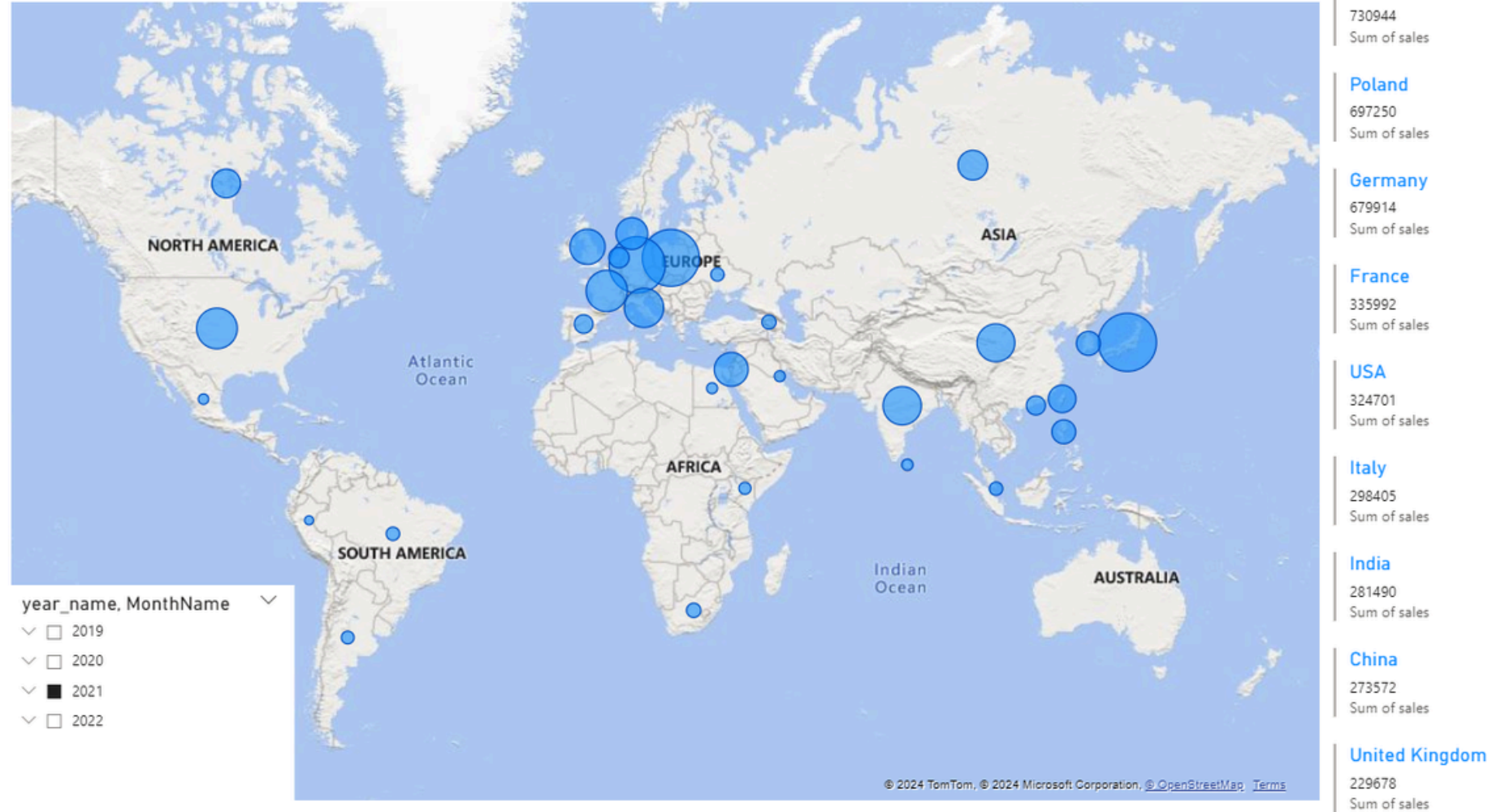


8. Data is loaded from the views to Power BI and data model relationships are checked.

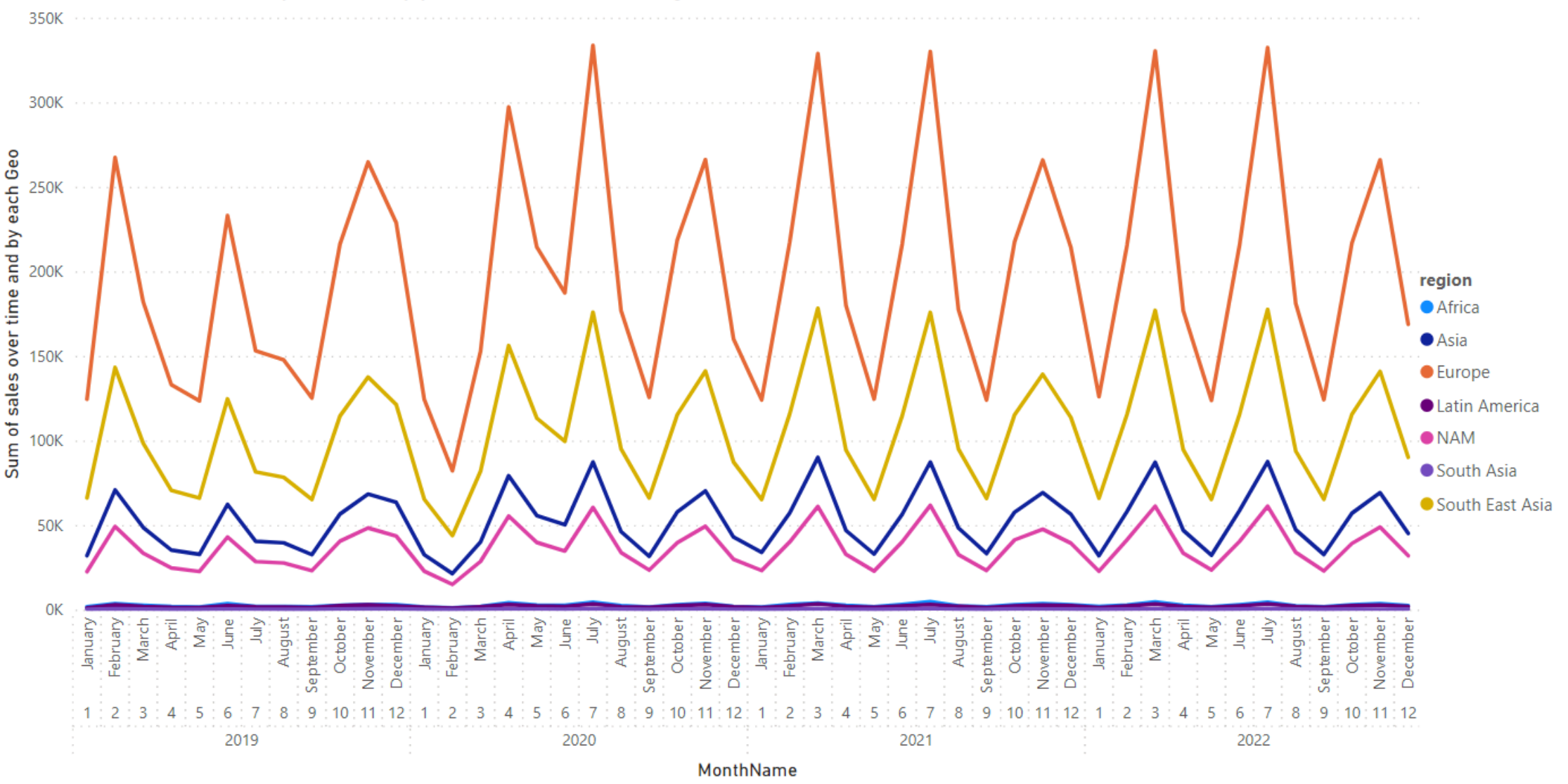


## 9. Reports building in Power BI

Sum of sales by country

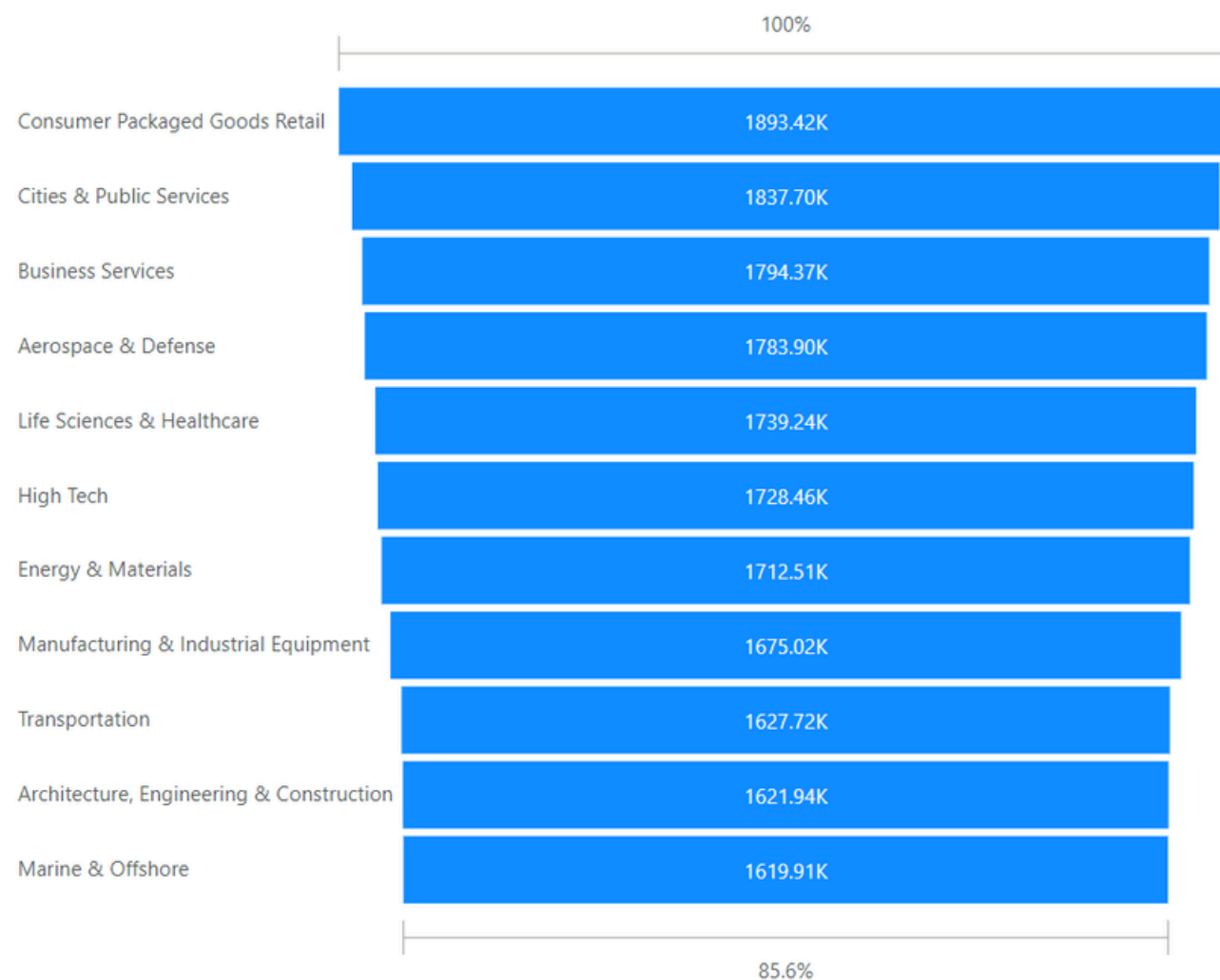


Sum of sales over time and by each Geo by year, /, MonthName and region





## Sum of sales by industry



## Apply Filters

country, state

- ☐ Argentina
- ☐ Armenia
- ☐ Brazil
- ☐ Canada
- ☐ China
- ☐ Denmark
- ☐ Egypt
- ☐ France
- ☐ Germany
- ☐ Hong Kong
- ☐ India
- ☐ Israel
- ☐ Italy
- ☐ Japan
- ☐ Kenya
- ☐ Kuwait
- ☐ Mexico
- ☐ Netherlands
- ☐ Peru
- ☐ Philippines
- ☐ Poland

year\_name, MonthName

- ☐ 2019
- ☐ 2020
- ☐ 2021
- ☐ 2022
- ☐ April
- ☐ August
- ☐ December
- ☐ February
- ☐ January
- ☐ July
- ☐ June
- ☐ March
- ☐ May
- ☐ November
- ☐ October
- ☐ September



## Information on Partner Companies and Partner Employees

28

Total Count of Partner Companies

289

Total Count of Partner Employees

partner_first_name	partner_last_name	partner_company	tgt_job_role	Target Job Roles approved for the partner
FirstName1199	LastName1199	Integrated Technologies	Manufacturing Engineer	1
FirstName1207	LastName1207	Future Innovations	Manufacturing Engineer	1
FirstName1233	LastName1233	Solutions Group	Manufacturing Engineer	1
FirstName1244	LastName1244	Rapid Prototyping Services	Manufacturing Engineer	1
FirstName1252	LastName1252	CAD Masters	Manufacturing Engineer	1
FirstName1256	LastName1256	Rapid Prototyping Services	Manufacturing Engineer	1
FirstName1260	LastName1260	Matrix Engineering	Manufacturing Engineer	1
Total				291

### Apply Filters

Filter Partner Company by Name

All

Filter Partner Job Role They are Allowed to Sell To

Manufacturing Engineer

partner_company	partner_first_name	partner_last_name	year	MonthName	license_type	Sum of sales
Advanced Dynamics	FirstName1010	LastName1010	2016	April	Premium	207
Advanced Dynamics	FirstName1010	LastName1010	2016	April	Professional	219
Advanced Dynamics	FirstName1010	LastName1010	2016	April	Standard	126
Advanced Dynamics	FirstName1010	LastName1010	2016	August	Premium	114
Advanced Dynamics	FirstName1010	LastName1010	2016	August	Professional	266
Advanced Dynamics	FirstName1010	LastName1010	2016	August	Standard	111
Advanced Dynamics	FirstName1010	LastName1010	2016	December	Premium	175
Advanced Dynamics	FirstName1010	LastName1010	2016	December	Professional	121
Advanced Dynamics	FirstName1010	LastName1010	2016	December	Standard	159
Advanced Dynamics	FirstName1010	LastName1010	2016	February	Premium	142
Total						1011439

Partner Company name	Total number of employees
Advanced Dynamics	9
Apex Systems	10
CAD Masters	13
Core Engineering Solutions	8
Dynamic Systems	10
Edge Technologies	11
Environmental Technologies	10
Flex Systems	11
Future Innovations	12
Horizon Engineering	6
Total	286