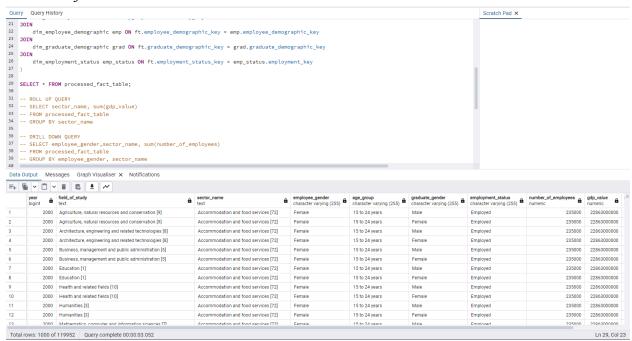
CSI4142

PHASE 3: Data Mining

Fact Table joined with all dimensions:



Rollup Query:

Query Query History 21 JOIN 22 dim_employee_demographic emp ON ft.employee_demographic_key = emp.employee_demographic_key 23 JOIN 24 dim_graduate_demographic grad ON ft.graduate_demographic_key = grad.graduate_demographic_key 25 JOIN 26 $\verb|dim_employment_status| emp_status| & \verb|oN| | ft.employment_status_key = emp_status.employment_key| \\$ 27) 28 29 -- SELECT * FROM processed_fact_table; 30 31 -- ROLL UP QUERY 32 SELECT sector_name, sum(gdp_value) 33 FROM processed_fact_table 34 GROUP BY sector_name 35 36 -- DRILL DOWN QUERY 37 -- SELECT employee_gender,sector_name, sum(number_of_employees) 38 -- FROM processed_fact_table 39 -- GROUP BY employee_gender, sector_name 40 Data Output Messages Graph Visualiser X Notifications **□ ~ □ ~ ■** sector_name sum numeric text 449604624000000 Information, culture and recreation [51, 71] 2 Agriculture, forestry, fishing and hunting [11] 217408464000000 3 Retail trade [44-45] 576352560000000 4 Utilities [22] 267908256000000 5 Public administration [91] 778441440000000 794953824000000 6 Construction [23] Accommodation and food services [72] 2392420800000000 8 731486784000000 Finance and insurance [52]

633725232000000

796667424000000

373839648000000

735803040000000

495811680000000

Total rows: 17 of 17 Query complete 00:00:00.922

Professional, scientific and technical services [54]

Business, building and other support services [55, 5...

Mining, quarrying, and oil and gas extraction [21]

Health care and social assistance [62]

Transportation and warehousing [48-49]

9

10

11

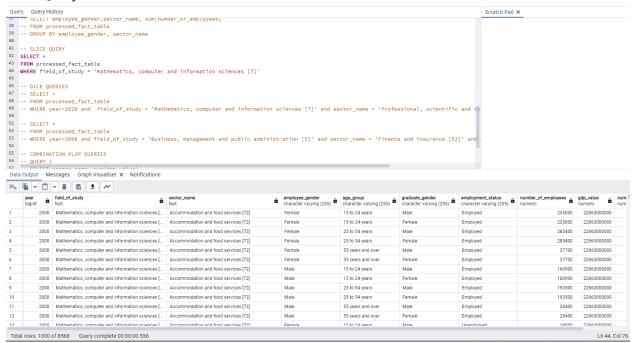
12

13

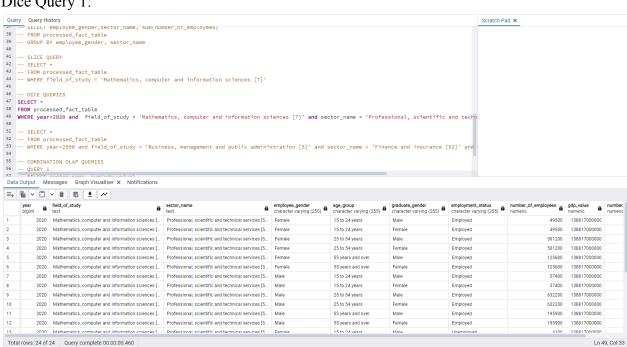
Drill down query:

```
Query Query History
31 -- ROLL UP QUERY
32 -- SELECT sector_name, sum(gdp_value)
33 -- FROM processed_fact_table
34 -- GROUP BY sector_name
35
36 -- DRILL DOWN QUERY
37 SELECT employee_gender,sector_name, sum(number_of_employees)
38 FROM processed_fact_table
39 GROUP BY employee_gender, sector_name
41 -- SLICE QUERY
42 -- SELECT *
43 -- FROM processed_fact_table
44 -- WHERE field_of_study = 'Mathematics, computer and information sciences [7]'
46 -- DICE QUERIES
47 -- SELECT *
48 -- FROM processed_fact_table
49 -- WHERE year=2020 and field_of_study = 'Mathematics, computer and information sciences [7]' and sector_name = 'Professional,
50
Data Output Messages Graph Visualiser X Notifications
=+ □ ∨ □ ∨ ■ ■ ★ ៷
      employee_gender character varying (255) a sector_name text
                                                                   a sum numeric
                                                                        405574400
      Female
                            Accommodation and food services [72]
      Female
                            Agriculture, forestry, fishing and hunting [11]
                                                                        65234008
      Female
                            Business, building and other support services [55, 5...
                                                                        195702472
4
      Female
                            Construction [23]
                                                                        87223668
                                                                        475834800
      Female
                            Educational services [61]
                            Finance and insurance [52]
                                                                        276902136
      Female
      Female
                            Health care and social assistance [62]
                                                                        984961600
                                                                        215257952
      Female
                            Information, culture and recreation [51, 71]
      Female
                            Manufacturing [31-33]
                                                                        340667600
10
      Female
                            Mining, quarrying, and oil and gas extraction [21]
                                                                        27851320
11
      Female
                            Other services (except public administration) [81]
                                                                        240538620
12
                            Professional, scientific and technical services [54]
                                                                        321383216
13
                                                                        267753780
Total rows: 34 of 34 Query complete 00:00:00.529
```

Slice Query:



Dice Query 1:



Dice Query 2:

Query Query History

77 --- Statch Pad X

78 --- FROW processed_fact_table
39 --- GROUP BY employee_gender, sector_name
40

41 --- SLICE QUERY
42 --- SELECT *
43 --- FROW processed_fact_table
44 --- WHERE field_of_study = 'Mathematics, computer and information sciences [7]'
45 --- DICE QUERIES
47 --- SELECT *
48 --- WHERE year=2008 and field_of_study = 'Mathematics, computer and information sciences [7]' and sector_name = 'Professional, scientific and to select the select of the select o

=+		v i								
	year bigint	field_of_study text	sector_name text	employee_gender character varying (255)	age_group character varying (255)	graduate_gender character varying (255)	employment_status character varying (255)	number_of_employees anumeric	gdp_value numeric	number_of_graduates numeric
1	2008	Business, management and public administration [Finance and insurance [52]	Female	15 to 24 years	Male	Employed	40200	99827000000	73371.0
2	2008	Business, management and public administration [Finance and insurance [52]	Female	15 to 24 years	Female	Employed	40200	99827000000	102933.0
3	2008	Business, management and public administration [Finance and insurance [52]	Female	25 to 54 years	Male	Employed	377300	99827000000	73371.0
4	2008	Business, management and public administration [Finance and insurance [52]	Female	25 to 54 years	Female	Employed	377300	99827000000	102933.0
5	2008	Business, management and public administration [Finance and insurance [52]	Female	55 years and over	Male	Employed	58100	99827000000	73371.0
6	2008	Business, management and public administration [Finance and insurance [52]	Female	55 years and over	Female	Employed	58100	99827000000	102933.0
7	2008	Business, management and public administration [Finance and insurance [52]	Male	15 to 24 years	Male	Employed	21000	99827000000	73371.0
8	2008	Business, management and public administration [Finance and insurance [52]	Male	15 to 24 years	Female	Employed	21000	99827000000	102933.0
9	2008	Business, management and public administration (Finance and insurance [52]	Male	25 to 54 years	Male	Employed	234800	99827000000	73371.0
10	2008	Business, management and public administration [Finance and insurance [52]	Male	25 to 54 years	Female	Employed	234800	99827000000	102933.0
11	2008	Business, management and public administration [Finance and insurance [52]	Male	55 years and over	Male	Employed	43500	99827000000	73371.0
12	2008	Business, management and public administration [Finance and insurance [52]	Male	55 years and over	Female	Employed	43500	99827000000	102933.0

Total rows: 12 of 12 | Query complete 00:00:00.462 | Ln 50, Col 1

Combined OLAP Query 1:

```
Query Query History
43 -- FROM processed_fact_table
44 -- WHERE field_of_study = 'Mathematics, computer and information sciences [7]'
45
46 -- DICE QUERIES
47 -- SELECT *
48 -- FROM processed_fact_table
49 -- WHERE year=2020 and field_of_study = 'Mathematics, computer and information sciences [7]' and sector_name
 50
51 -- SELECT *
52 -- FROM processed_fact_table
 53 -- WHERE year=2008 and field_of_study = 'Business, management and public administration [5]' and sector_name =
 54
55 -- COMBINATION OLAP QUERIES
56 -- QUERY 1
57 SELECT sector_name, sum(gdp_value)
58 FROM processed_fact_table
 59 WHERE year = 2008
60 GROUP BY sector_name
 61
 Data Output Messages Graph Visualiser X Notifications
□ ■ * *
                                            sum
      sector_name
                                                numeric
                                                 20815872000000
      Information, culture and recreation [51, 71]
 2
                                                  9845472000000
      Agriculture, forestry, fishing and hunting [11]
 3
      Retail trade [44-45]
                                                 26958960000000
 4
      Utilities [22]
                                                 12635280000000
      Public administration [91]
                                                 33878880000000
 6
                                                 37636368000000
      Construction [23]
      Accommodation and food services [72]
                                                 10668672000000
      Finance and insurance [52]
                                                 33541872000000
 8
                                                 27838272000000
      Professional, scientific and technical services [54]
                                                 34405056000000
 10
      Health care and social assistance [62]
 11
      Business, building and other support services [55, 5...
                                                 17958864000000
 12
      Mining, quarrying, and oil and gas extraction [21]
                                                 52687152000000
 13
      Transportation and warehousing [48-49]
                                                 212520000000000
Total rows: 17 of 17 Query complete 00:00:00.773
```

Combined OLAP Query 2:

```
Query Query History
21
    JOIN
22
         dim_employee_demographic emp ON ft.employee_demographic_key = emp.employee_demographic_key
23
    JOIN
24
         \verb|dim_graduate_demographic grad ON ft.graduate_demographic_key = grad.graduate_demographic_key| \\
25
    JOIN
26
         dim_employment_status emp_status ON ft.employment_status_key = emp_status.employment_key
27
28
29 -- QUERY 2
30 drill_down_table AS (
31
      SELECT year, employee_gender, sector_name, sum(number_of_employees)
32
      FROM processed_fact_table
33
      GROUP BY year,employee_gender, sector_name
34
35
   SELECT employee_gender, sector_name, sum as number_of_employees
36
    FROM drill_down_table
37
    WHERE drill_down_table.year = 2008
38
39 -- QUERY 3
40 -- drill_down_table AS (
Data Output Messages Graph Visualiser X Notifications
=+
    number_of_employees
      employee_gender
                            sector_name
      character varying (255)
                            text
                                                                       numeric
      Female
                            Accommodation and food services [72]
                                                                                   19611200
2
      Female
                            Agriculture, forestry, fishing and hunting [11]
                                                                                    3137540
3
      Female
                            Business, building and other support services [55, 5...
                                                                                    9926000
4
      Female
                            Construction [23]
                                                                                    4390400
5
                            Educational services [61]
                                                                                   22052800
      Female
6
                            Finance and insurance [52]
                                                                                   13562080
      Female
7
      Female
                            Health care and social assistance [62]
                                                                                   44102800
8
                            Information, culture and recreation [51, 71]
      Female
                                                                                   10231200
9
                            Manufacturing [31-33]
                                                                                   16287600
      Female
10
      Female
                            Mining, quarrying, and oil and gas extraction [21]
                                                                                    1492400
                            Other services (except public administration) [81]
                                                                                   11737600
11
      Female
12
      Female
                            Professional, scientific and technical services [54]
                                                                                   14462000
13
      Female
                            Public administration [91]
                                                                                   13104000
Total rows: 34 of 34 Query complete 00:00:00.753
```

Combined OLAP Query 3:

```
Query Query History
34 __ )
35 -- SELECT employee_gender, sector_name, sum as number_of_employees
36 -- FROM drill_down_table
37 -- WHERE drill_down_table.year = 2008
38
39 -- QUERY 3
40 drill_down_table AS (
41
      SELECT year, graduate_gender, field_of_study, sum(number_of_graduates)
42
      FROM processed_fact_table
43
      GROUP BY year, graduate_gender, field_of_study
44
45 SELECT graduate_gender, field_of_study, sum as number_of_graduates
46 FROM drill_down_table
47 WHERE drill_down_table.year = 2008
48
49 -- QUERY 4
50 -- slice_table AS (
51 -- SELECT *
52 -- FROM processed_fact_table
53 -- WHERE year=2008
Data Output Messages Graph Visualiser X Notifications
=+ | • | • | • | • |
                              ! ~
                                                                                 number_of_graduates
                            field_of_study
      graduate_gender
      character varying (255)
                                                                                 numeric
                                                                                            1379448.0
      Female
                             Agriculture, natural resources and conservation [9]
2
                                                                                            3147516.0
      Female
                             Architecture, engineering and related technologies [8]
3
      Female
                             Business, management and public administration [5]
                                                                                           20998332.0
4
                                                                                            9957240.0
      Female
                             Education [1]
                             Health and related fields [10]
                                                                                           17664768.0
5
      Female
      Female
                                                                                           11995200.0
6
                             Humanities [3]
7
                             Mathematics, computer and information sciences [7]
                                                                                            1481652.0
      Female
8
      Female
                             Other [12]
                                                                                             477972.0
9
      Female
                             Personal improvement and leisure [0]
                                                                                             343332.0
10
      Female
                             Personal, protective and transportation services [11]
                                                                                            1899036.0
11
      Female
                             Physical and life sciences and technologies [6]
                                                                                            6320736.0
                             Social and behavioural sciences and law [4]
                                                                                           18143352.0
12
      Female
13
      Female
                             Unclassified
                                                                                               7344.0
Total rows: 28 of 28
                      Query complete 00:00:00.611
```

Combined OLAP Query 4:

```
Query Query History
39 -- QUERY 3
40 -- drill_down_table AS (
41 -- SELECT year, graduate_gender, field_of_study, sum(number_of_graduates)
42 -- FROM processed_fact_table
43
        GROUP BY year, graduate_gender, field_of_study
44 -- )
45 -- SELECT graduate_gender, field_of_study, sum as number_of_graduates
46
    -- FROM drill_down_table
47 -- WHERE drill_down_table.year = 2008
48
49 -- QUERY 4
50 slice_table AS (
51
         SELECT *
52
         FROM processed_fact_table
53
         WHERE year=2008
54 )
55 SELECT sector_name, employment_status, sum(number_of_employees) as number_of_people
56 FROM slice_table
57 GROUP BY sector_name,employment_status
ORDER BY sector_name, employment_status
Data Output Messages Graph Visualiser X Notifications
                              =+ □ ∨ □ ∨
                    number_of_people
      sector_name
                                                 employment_status
                                                character varying (255)
      text
1
      Accommodation and food services [72]
                                                 Employed
                                                                               30466800
2
      Accommodation and food services [72]
                                                 Unemployed
                                                                                2340800
3
      Agriculture, forestry, fishing and hunting [11]
                                                 Employed
                                                                               11074000
4
      Agriculture, forestry, fishing and hunting [11]
                                                 Unemployed
                                                                                 824740
5
      Business, building and other support services [55, 5...
                                                 Employed
                                                                               19843600
      Business, building and other support services [55, 5...
                                                 Unemployed
                                                                                1870400
6
7
      Construction [23]
                                                 Employed
                                                                               35072800
                                                                                2648800
8
      Construction [23]
                                                 Unemployed
9
      Educational services [61]
                                                                               32116000
                                                 Employed
10
      Educational services [61]
                                                 Unemployed
                                                                                1145200
11
      Finance and insurance [52]
                                                 Employed
                                                                               21697200
12
      Finance and insurance [52]
                                                 Unemployed
                                                                                 446880
13
      Health care and social assistance [62]
                                                 Employed
                                                                               52869600
Total rows: 34 of 34 Query complete 00:00:00.611
```

Iceberg query:

Query Query History 1 -- ICEBERG QUERY 2 WITH gdp_per_industry AS (SELECT 3 dt.year, 4 ind.sector_name, 5 ft.gdp_value FROM fact_table ft 6 JOIN 7 dim_date dt ON ft.date_key = dt.date_key 8 JOIN 9 dim_industry ind ON ft.industry_key = ind.industry_key) 10 11 SELECT DISTINCT * 12 FROM gdp_per_industry 13 ORDER BY gdp_value DESC 14 LIMIT 10 15 16 17 -- WINDOWING QUERY 18 -- Total # of grads per field of study per gender between 2015 and 2020 19 20 -- WITH graduates_per_field_of_study AS (Data Output Messages Graph Visualiser X Notifications =+ sector_name gdp_value year â bigint text numeric 1 2020 Real estate and rental and leasing [53] 293886000000 2 2019 Real estate and rental and leasing [53] 282688000000 3 2018 Real estate and rental and leasing [53] 267102000000 4 2017 Real estate and rental and leasing [53] 257949000000 5 2016 Real estate and rental and leasing [53] 249016000000 2015 Real estate and rental and leasing [53] 238235000000 6 7 2014 Real estate and rental and leasing [53] 228526000000 8 2018 Manufacturing [31-33] 220573000000 9 2013 Real estate and rental and leasing [53] 219018000000

215153000000

Manufacturing [31-33]

10

2019

Windowing query:

```
Query Query History
37
    FROM
38
         (SELECT field_of_study, graduate_gender, sum(sum1) as total_graduates
39
              FROM (SELECT year, field_of_study, graduate_gender, sum(number_of_graduates) as sum1
40
                       FROM graduates_per_field_of_study
41
                       WHERE year >= 2015 AND year <= 2020
42
                       GROUP BY year, field_of_study, graduate_gender) AS total_graduates_per_field_of_study
43
             GROUP BY field_of_study, graduate_gender
44
         ) AS rankings_fields_of_study
45
46
47 -- USING THE WINDOW CLAUSE
48
    -- WITH gdp_per_industry AS (SELECT
49 __
           dt.year,
50 __
            ind.sector_name,
51 __
            ft.gdp_value FROM fact_table ft
52 -- JOIN
53 __
           dim_date dt ON ft.date_key = dt.date_key
54 -- JOIN
55 __
            dim_industry ind ON ft.industry_key = ind.industry_key)
56
Data Output Messages Graph Visualiser X Notifications
₹ ~
                                                                                              field_rank
                                                                              total_graduates
      field_of_study
                                                         graduate_gender
                                                         character varying (255)
                                                                              numeric
                                                         Female
                                                                                   176748048.0
                                                                                                        1
      Business, management and public administration [5]
2
                                                         Female
                                                                                   155810304.0
                                                                                                        2
      Health and related fields [10]
3
      Architecture, engineering and related technologies [8]
                                                         Male
                                                                                   146087460.0
                                                                                                        3
                                                         Male
4
      Business, management and public administration [5]
                                                                                   143909964.0
                                                                                                        4
                                                                                   143762472.0
                                                                                                        5
5
      Social and behavioural sciences and law [4]
                                                         Female
      Humanities [3]
                                                         Female
                                                                                   69838992.0
                                                                                                        6
6
      Education [1]
                                                         Female
                                                                                    55200564.0
                                                                                                        7
8
      Social and behavioural sciences and law [4]
                                                         Male
                                                                                    54968004.0
                                                                                                        8
                                                                                    47865744.0
                                                                                                        9
9
      Physical and life sciences and technologies [6]
                                                         Female
10
      Health and related fields [10]
                                                         Male
                                                                                    42226164.0
                                                                                                       10
11
      Humanities [3]
                                                         Male
                                                                                    40314276.0
                                                                                                       11
```

Male

Male

38291004.0

35836272.0

12

13

13 Physical and life sciences and technologies [6]

Total rows: 28 of 28 Query complete 00:00:00.386

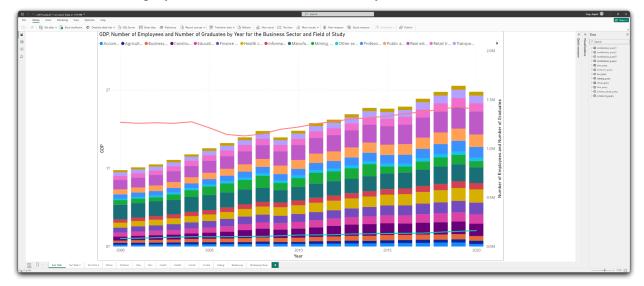
Mathematics, computer and information sciences [7]

12

Window clause query:

```
Query History
Query
         ) AS rankings_fields_of_study
45
46
47
   -- USING THE WINDOW CLAUSE
48
    WITH gdp_per_industry AS (SELECT
49
         dt.year,
50
         ind.sector_name,
51
         ft.gdp_value FROM fact_table ft
52 JOIN
53
         dim_date dt ON ft.date_key = dt.date_key
54
    JOIN
55
         dim_industry ind ON ft.industry_key = ind.industry_key)
56
57 SELECT DISTINCT year, sector_name, avg(gdp_value) OVER W moving_average
58
    FROM gdp_per_industry
59
    WINDOW W AS (PARTITION BY sector_name
60
                         ORDER BY year
61
                         RANGE BETWEEN 1 PRECEDING AND 1 FOLLOWING
62
63 ORDER BY year ASC
Data Output
             Messages Graph Visualiser X Notifications
                                                                moving_average
      year
                  sector_name
                                                                                    â
      bigint
                                                                numeric
                                                                 123690500000.00000000
1
            2000
                  Real estate and rental and leasing [53]
2
            2000
                  Agriculture, forestry, fishing and hunting [11]
                                                                 23673000000.000000000
                                                                 58049500000.00000000
3
            2000
                  Mining, quarrying, and oil and gas extraction [21]
                                                                 47365000000.00000000
4
            2000
                  Transportation and warehousing [48-49]
5
            2000
                  Accommodation and food services [72]
                                                                 23220500000.00000000
            2000
                  Business, building and other support services [55, 5...
                                                                 30610000000.00000000
6
7
            2000
                   Construction [23]
                                                                 53308500000.00000000
8
            2000
                  Educational services [61]
                                                                  53136500000.00000000
9
            2000
                  Utilities [22]
                                                                 26956500000.00000000
10
            2000
                  Professional, scientific and technical services [54]
                                                                 53310000000.00000000
11
            2000
                  Other services (except public administration) [81]
                                                                 21774500000.000000000
                  Health care and social assistance [62]
12
            2000
                                                                 65502000000.00000000
13
            2000
                  Public administration [91]
                                                                 66005500000.00000000
Total rows: 357 of 357
                         Query complete 00:00:00.479
```

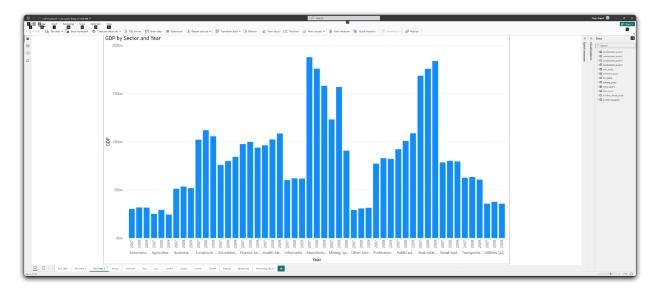
GDP, Number of Employees and Number of Graduates by Year for the Business Sector



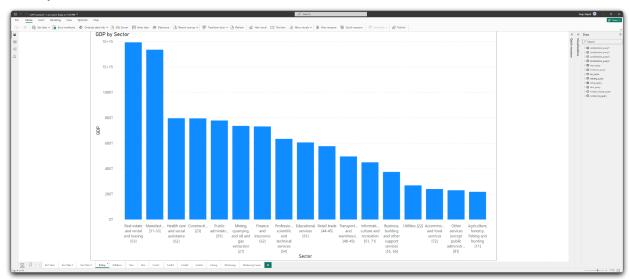
Number of Employees and Number of Graduates by Year



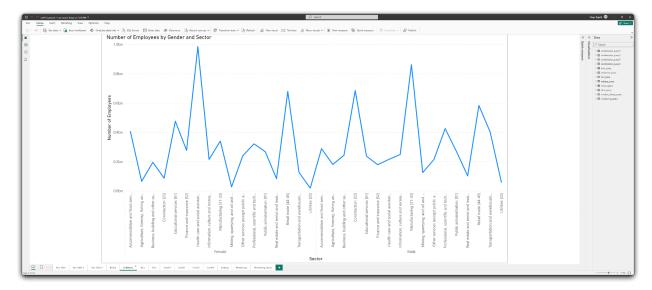
GDP by Sector and Year



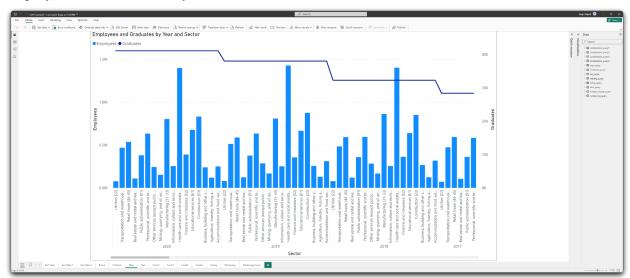
GDP by Sector



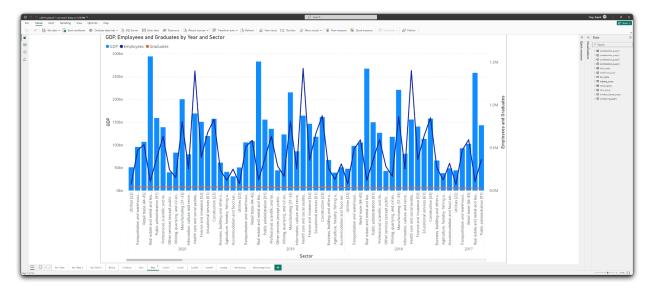
Number of Employees by Gender and Sector



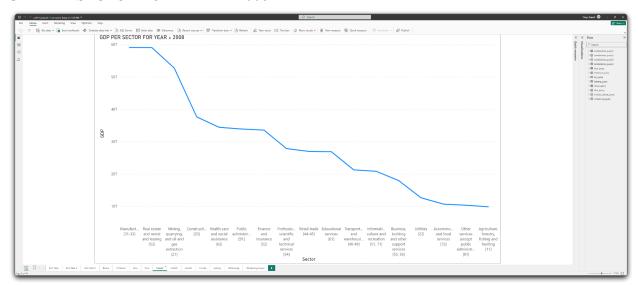
Employees and Graduates by Year and Sector



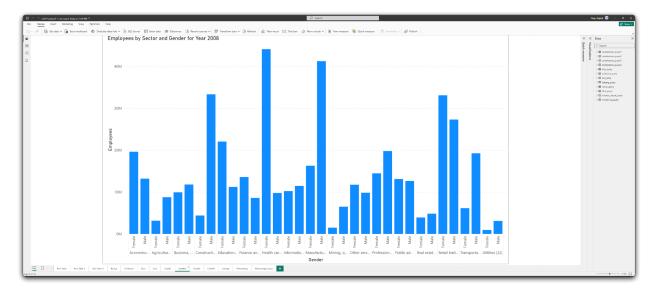
GDP, Employees and Graduates by Year and Sector



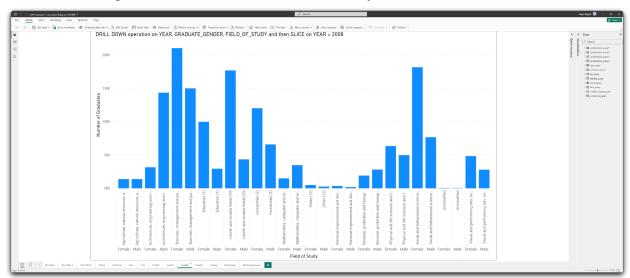
GDP PER SECTOR FOR YEAR = 2008



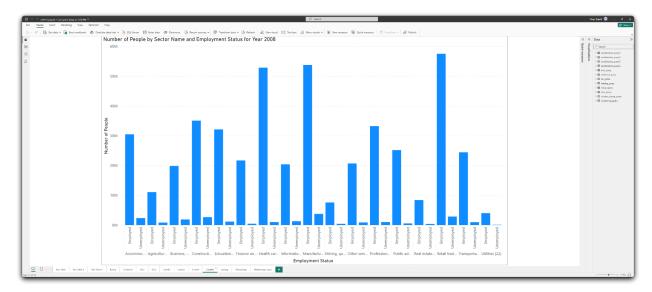
Employees by Sector and Gender for Year 2008



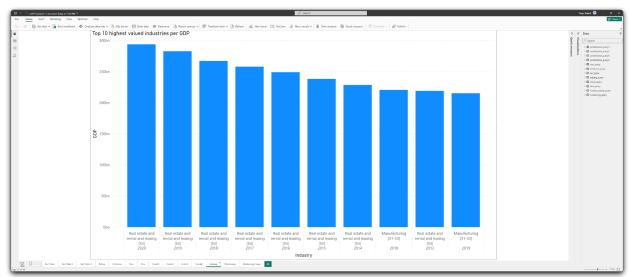
DRILL DOWN operation on Year, Gender, Field of Study and then Slice on Year = 2008



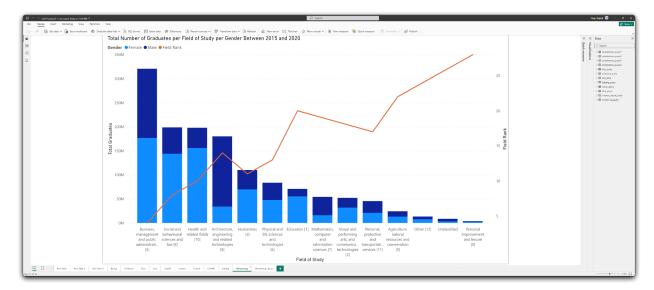
Number of People By Sector Name and Employment Status for Year 2008



Top 10 Highest Valued Industries per GDP



Total Number of Graduates per Field of Study per Gender Between 2015 and 2020



Moving Average of GDP per Industry over Three Years

