- Review IDMA Chapter 6 and author slide presentation
    - Pages 211-252 and 268-272

Reviewed IDMA Chapter 6 and author slide presentation

- For each of the data items in the corgis-covid.csv dataset, identify its analytical data type

Date. Day - Ordinal

Date. Month - Ordinal

Date. Year - Ordinal (can be Nominal (categorical) – as it contains only 2020 year, no math can be done in this dataset using this year. So it can be nominal too)

Data. Cases – Ratio

Data. Deaths - Ratio

Location. Country - Nominal

Location. Code - Nominal

Data. Population - Ratio

Location. Continent - Nominal

Data. Rate – Ratio

- Create a SQL database and table for the dataset using a RDBMS (Oracle, MySQL, etc.)

Created a table called COVID_DATA in oracle 18 c Database.

- Load the dataset into the table; use SQL to display a few records

Loaded the dataset into the table with the help of Import Data function in oracle SQL Developer.

Display of few records:

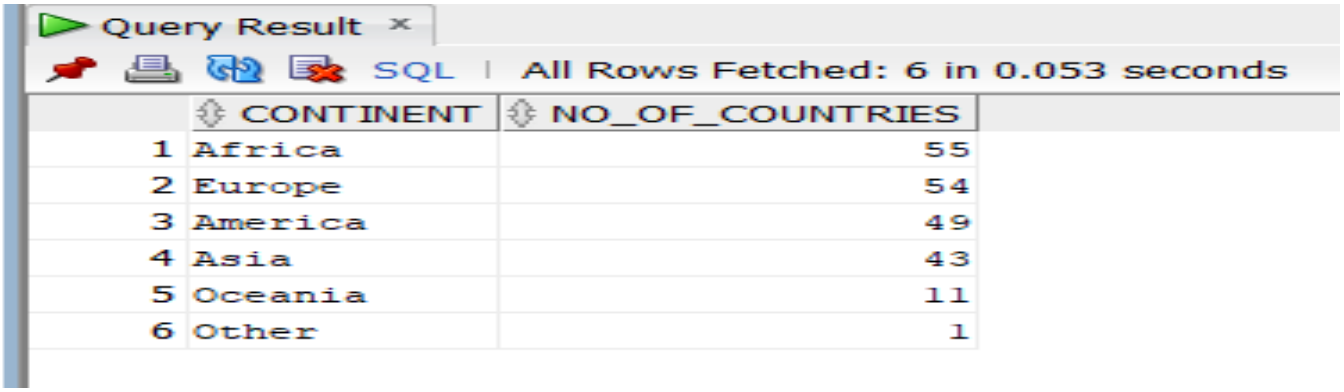
Select * from covid_data where rownum <= 5

	DAY	MONTH	YEAR	CASES	DEATHS	COUNTRY	CODE	POPULATION	CONTINENT	RATE
1	26	6	2020	460	36	Afghanistan	AFG	38041757	Asia	19.15000929
2	25	6	2020	234	21	Afghanistan	AFG	38041757	Asia	19.90444343
3	24	6	2020	338	20	Afghanistan	AFG	38041757	Asia	21.08735409
4	23	6	2020	310	17	Afghanistan	AFG	38041757	Asia	21.62360692
5	22	6	2020	409	12	Afghanistan	AFG	38041757	Asia	22.32020987

- Query the database and interpret the results, displaying:

    - the number of countries in each continent

Select continent, count(distinct(country)) as No_of_Countries from covid_data group by continent order by No_of_Countries desc

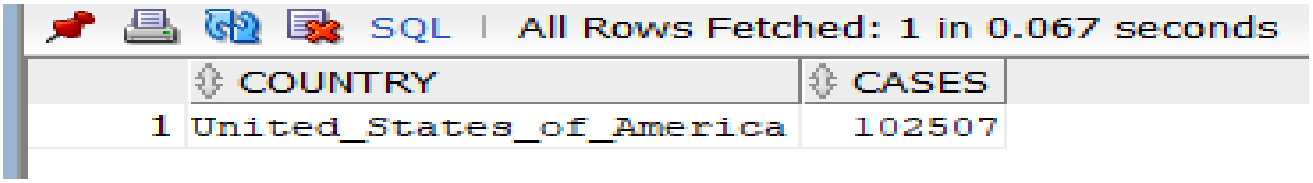

Query Result

SQL | All Rows Fetched: 6 in 0.053 seconds

	CONTINENT	NO_OF_COUNTRIES
1	Africa	55
2	Europe	54
3	America	49
4	Asia	43
5	Oceania	11
6	Other	1

    - the country with the highest number of new daily cases

select country,cases from covid_data where cases=(select max(cases) from covid_data)

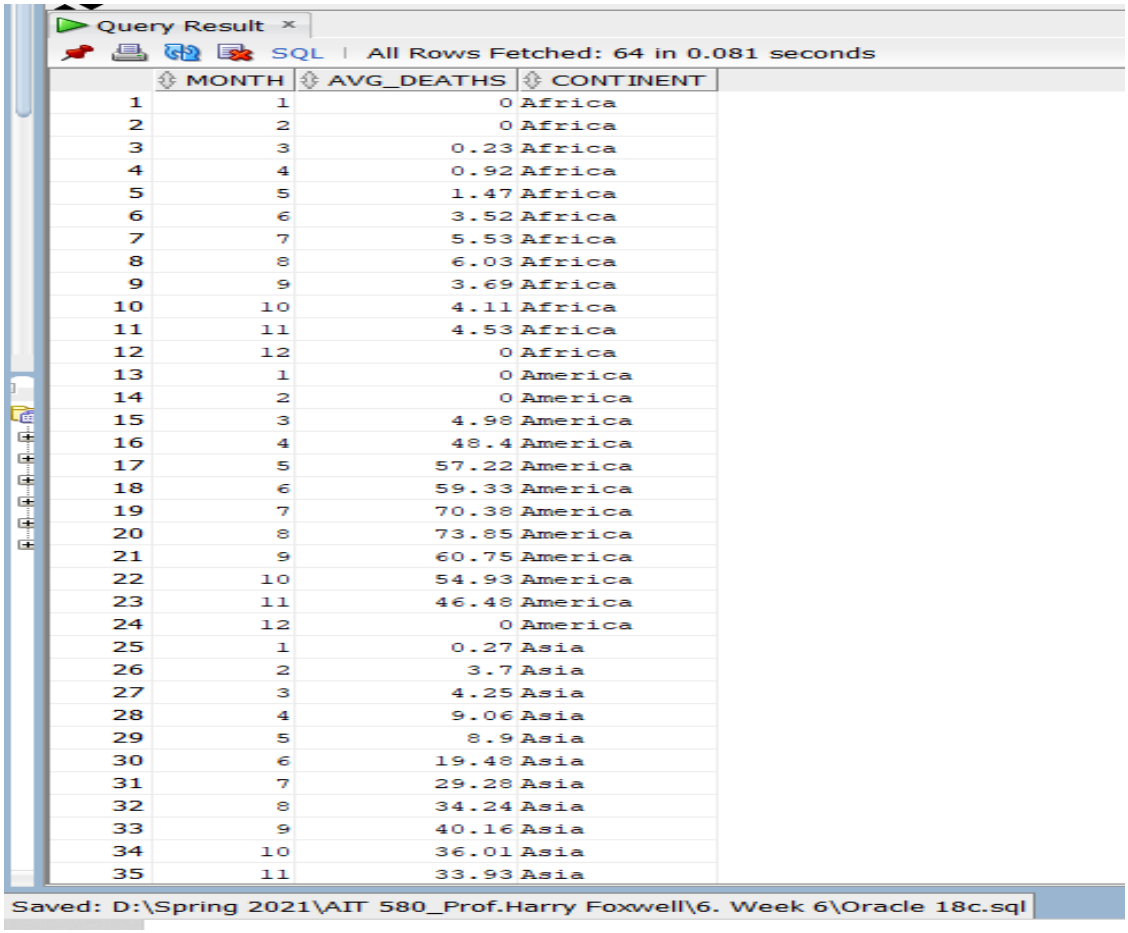

SQL | All Rows Fetched: 1 in 0.067 seconds

	COUNTRY	CASES
1	United_States_of_America	102507

United States of America has the highest number of new daily cases.

    - the mean number of monthly deaths for each continent

select month,round(avg(deaths),2) as Avg_Deaths,continent from covid_data group by month,continent order by continent

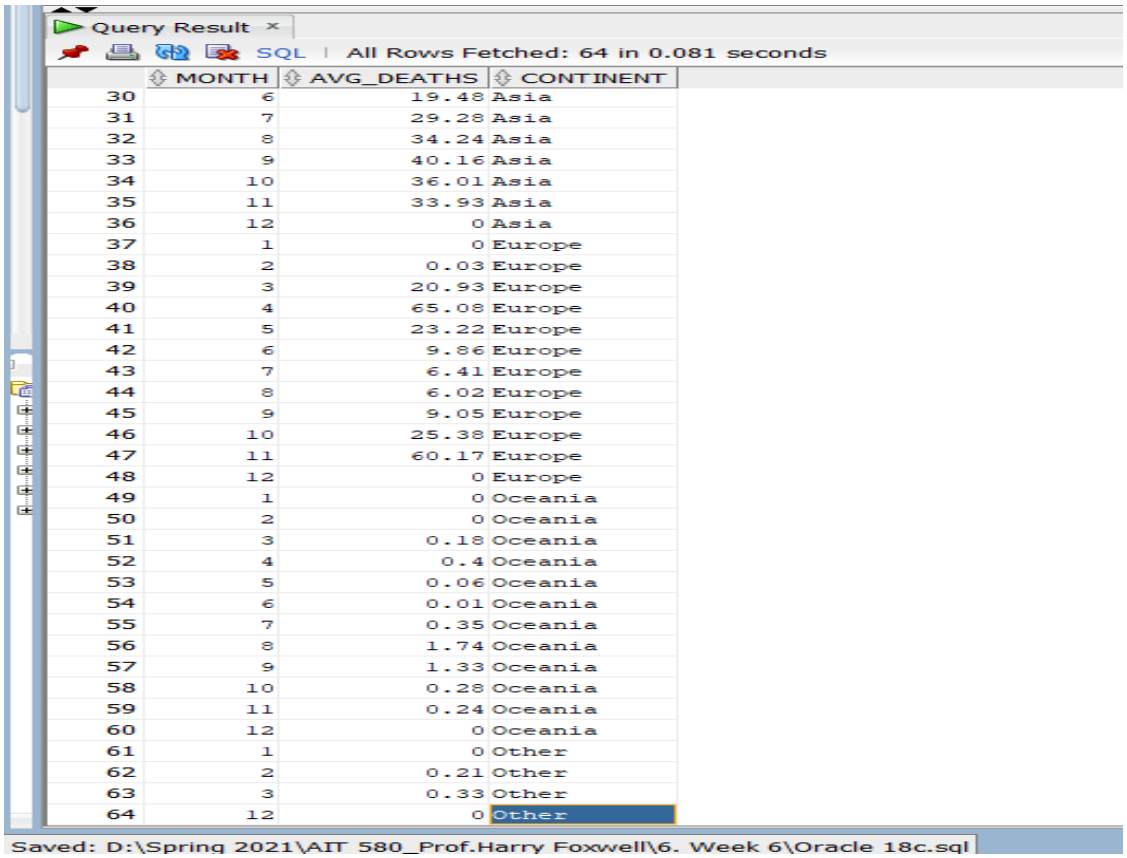


Query Result

SQL | All Rows Fetched: 64 in 0.081 seconds

	MONTH	AVG_DEATHS	CONTINENT
1	1	0	Africa
2	2	0	Africa
3	3	0.23	Africa
4	4	0.92	Africa
5	5	1.47	Africa
6	6	3.52	Africa
7	7	5.53	Africa
8	8	6.03	Africa
9	9	3.69	Africa
10	10	4.11	Africa
11	11	4.53	Africa
12	12	0	Africa
13	1	0	America
14	2	0	America
15	3	4.98	America
16	4	48.4	America
17	5	57.22	America
18	6	59.33	America
19	7	70.38	America
20	8	73.85	America
21	9	60.75	America
22	10	54.93	America
23	11	46.48	America
24	12	0	America
25	1	0.27	Asia
26	2	3.7	Asia
27	3	4.25	Asia
28	4	9.06	Asia
29	5	8.9	Asia
30	6	19.48	Asia
31	7	29.28	Asia
32	8	34.24	Asia
33	9	40.16	Asia
34	10	36.01	Asia
35	11	33.93	Asia

Saved: D:\Spring 2021\AIT 580_Prof.Harry Foxwell\6. Week 6\Oracle 18c.sql



Query Result

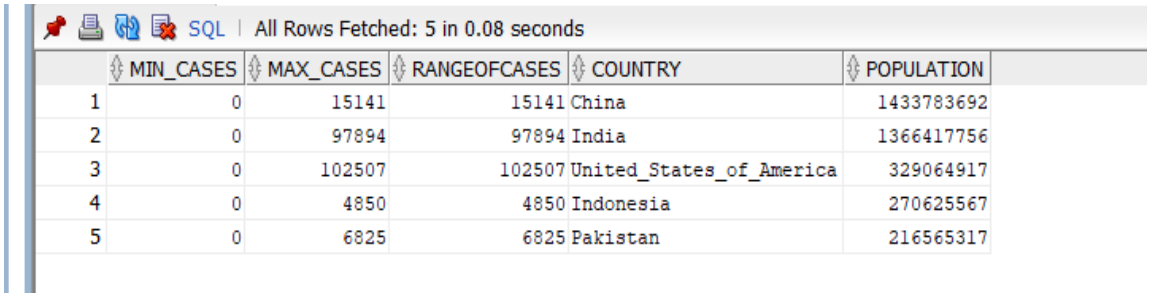
SQL | All Rows Fetched: 64 in 0.081 seconds

	MONTH	AVG_DEATHS	CONTINENT
30	6	19.48	Asia
31	7	29.28	Asia
32	8	34.24	Asia
33	9	40.16	Asia
34	10	36.01	Asia
35	11	33.93	Asia
36	12	0	Asia
37	1	0	Europe
38	2	0.03	Europe
39	3	20.93	Europe
40	4	65.08	Europe
41	5	23.22	Europe
42	6	9.86	Europe
43	7	6.41	Europe
44	8	6.02	Europe
45	9	9.05	Europe
46	10	25.38	Europe
47	11	60.17	Europe
48	12	0	Europe
49	1	0	Oceania
50	2	0	Oceania
51	3	0.18	Oceania
52	4	0.4	Oceania
53	5	0.06	Oceania
54	6	0.01	Oceania
55	7	0.35	Oceania
56	8	1.74	Oceania
57	9	1.33	Oceania
58	10	0.28	Oceania
59	11	0.24	Oceania
60	12	0	Oceania
61	1	0	Other
62	2	0.21	Other
63	3	0.33	Other
64	12	0	Other

Saved: D:\Spring 2021\AIT 580_Prof.Harry Foxwell\6. Week 6\Oracle 18c.sql

- the range of cumulative cases for the 5 most populous countries

select min(cases) as MIN_CASES ,max(cases) AS Max_CASES, max(cases)-min(cases) as rangeofcases, country, population from covid_data group by population, country order by population desc fetch first 5 rows only



SQL | All Rows Fetched: 5 in 0.08 seconds

	MIN_CASES	MAX_CASES	RANGEOFCASES	COUNTRY	POPULATION
1	0	15141	15141	China	1433783692
2	0	97894	97894	India	1366417756
3	0	102507	102507	United_States_of_America	329064917
4	0	4850	4850	Indonesia	270625567
5	0	6825	6825	Pakistan	216565317

- Review IDMA Chapter 6 and author slide presentation
  - Pages 291-315

Reviewed IDMA Chapter 6 and author slide presentation

- Create a Mongo database, load the corgis-covid.csv dataset into it.

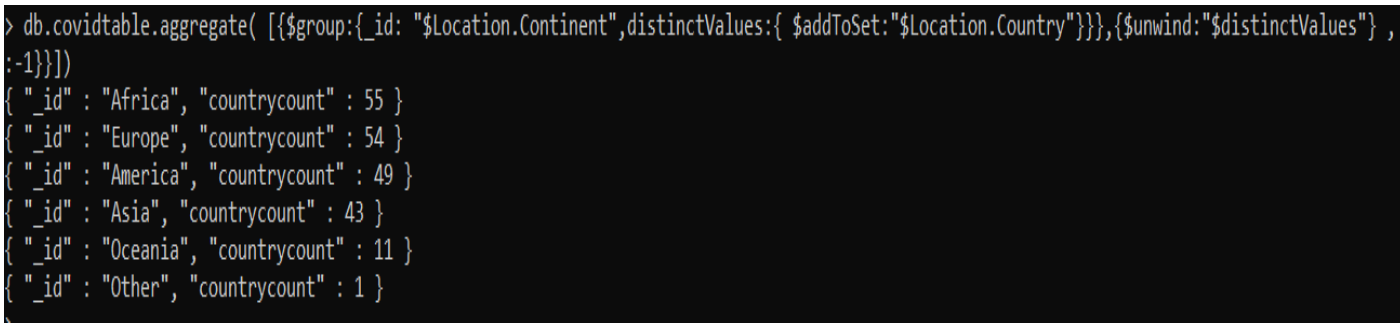
Created a Mongo Database named coviddatabase and Collection named covidtable.

Loaded the dataset:

mongoimport -d coviddatabase -c covidtable --type csv --file corgis-covid.csv –headerline

- Query the database, displaying:
  - the number of countries in each continent

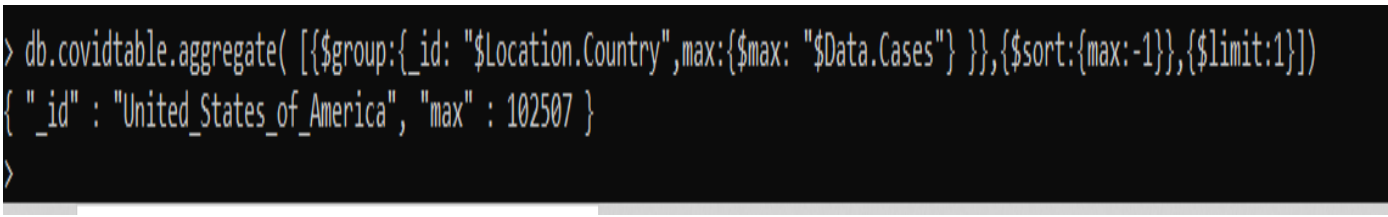
db.covidtable.aggregate([{$group:{_id: "$Location.Continent",distinctValues:{ $addToSet:"$Location.Country"}}},{$unwind:"$distinctValues"} , {$group:{_id:"$_id",countrycount:{$sum:1}}},{$sort:{countrycount:-1}}])



```
> db.covidtable.aggregate( [{$group:{_id: "$Location.Continent",distinctValues:{ $addToSet:"$Location.Country"}}},{$unwind:"$distinctValues"} ,
:-1}}])
{ "_id" : "Africa", "countrycount" : 55 }
{ "_id" : "Europe", "countrycount" : 54 }
{ "_id" : "America", "countrycount" : 49 }
{ "_id" : "Asia", "countrycount" : 43 }
{ "_id" : "Oceania", "countrycount" : 11 }
{ "_id" : "Other", "countrycount" : 1 }
```

- the country with the highest number of new daily cases

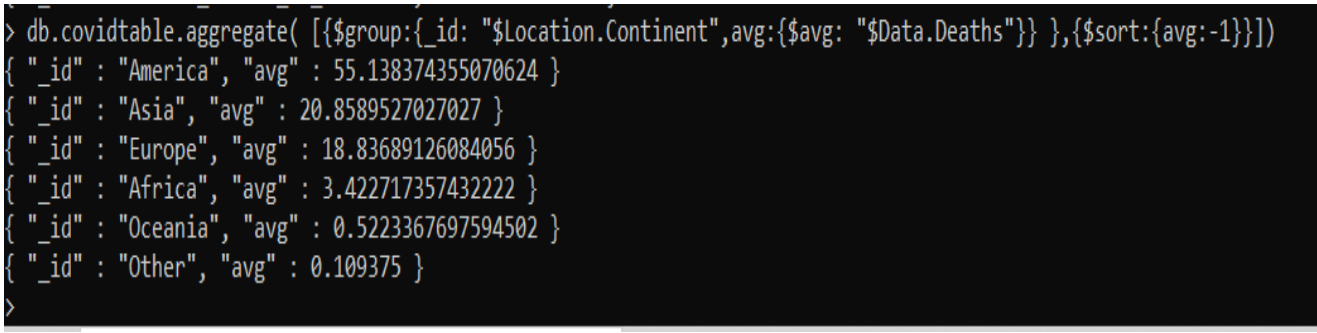
db.covidtable.aggregate([{$group:{_id: "$Location.Country",max:{$max: "$Data.Cases"} }},{$sort:{max:-1}},{$limit:1}])



United States of America has the highest number of new daily cases.

- the mean number of deaths per continent

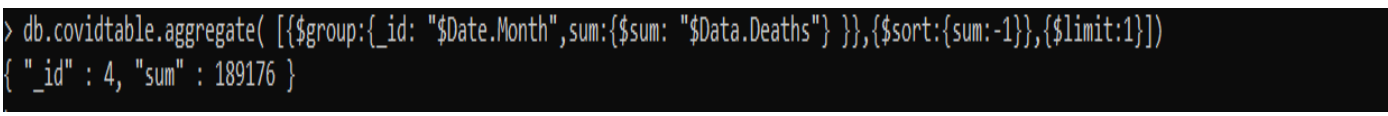
db.covidtable.aggregate([{$group:{_id: "$Location.Continent",avg:{$avg: "$Data.Deaths"}} },{$sort:{avg:-1}}])



America has the highest mean number of deaths.

- the month with the largest number of total deaths

db.covidtable.aggregate([{$group:{_id: "$Date.Month",sum:{$sum: "$Data.Deaths"} }},{$sort:{sum:-1}},{$limit:1}])



April had the largest number of total deaths of 189176.