Phase 3: OLAP Queries and BI Dashboard

CSI4142: Project

Part A: OLAP Queries

The queries can be found in the GitHub repository under \queries:

https://github.com/Daniel-Dumitrescu/csi4142-project.git

Part 1. Standard OLAP operations

Drill down and roll up [2 queries]

Query 1:

```
/*Roll up from month to quarter*/
SELECT D."Year",D."Quarter", SUM("Price") as Quarterly_Price , SUM("Volume") as
Quarterly_Volume
FROM public."FactTable" F, public."Date" D
WHERE F."Date_ID" = D."Date_ID"
GROUP BY D."Year", D."Quarter"
ORDER BY D."Year", D."Quarter" asc
```

Output for Query 1:

Year	Quarter	quarterly_price	quarterly_volume
2001	1	11602.97	4819287.7579606
2001	2	12457.57	3510166.7579606
2001	3	14235.66	3622939.58598687
2001	4	15473.76	3335835
2002	1	16782.87	2398792
2002	2	16669.52	2209206
2002	3	18364.21	3193139
2002	4	19652.71	3322360

Query 2:

```
/*Roll up provinces to Canada Total price and volume every month*/
SELECT F."Date_ID", SUM("Price") as Total_Price , SUM("Volume") as Total_Volume
FROM public."FactTable" F
GROUP BY F."Date_ID"
ORDER BY F."Date_ID" asc
```

Output for Query 2:

Date_ID	total_price	total_volume
0	3778.94	1615168.58598687
1	3852.21	1392225.58598687
2	3971.82	1811893.58598687
3	4033.36	1087495.58598687
4	4174.84	1229354.58598687
5	4249.37	1193316.58598687
6	4653	1311950.58598687
7	4790.24	939242

Slice [1 query]

Query 1:

```
/*Slice by Ontario*/
WITH region_id AS (
    SELECT * FROM public."Region" WHERE "Name_EN" = 'Ontario'
)
SELECT F.*
FROM public."FactTable" F, region_id
WHERE F."Region_ID" = region_id."Region_ID";
```

Output for Query 1:

Date_ID	Crop_ID	Climate_ID	Region_ID	Price	Volume	Total_Value
0	2	4	7	105	1761	184905
1	2	10	7	115.82	1761	203959.02
2	2	16	7	120.74	1761	212623.14
3	2	22	7	126.76	1761	223224.36
4	2	28	7	139.69	1761	245994.09
5	2	34	7	136.7	1761	240728.7
6	2	40	7	125.23	1761	220530.03
7	2	46	7	150.65	1761	265294.65

Dice [2 queries]

Query 1:

```
/*Dice by Alberta and 2019*/
WITH region_id AS (
    SELECT * FROM public."Region" WHERE "Name_EN" = 'Alberta'
), date_id AS (
    SELECT * FROM public."Date" WHERE "Year" = 2019
)
SELECT F.*
FROM public."FactTable" F, region_id, date_id
WHERE F."Region_ID" = region_id."Region_ID" AND F."Date_ID" = date_id."Date_ID";
```

Output for Query 1:

Date_ID	Crop_ID	Climate_ID	Region_ID	Price	Volume	Total_Value
216	0	1297	12	249.93	802296	200517839.28
216	1	1297	12	239.4	54336	13008038.4
216	2	1297	12	207.61	28550	5927265.5
216	3	1297	12	225.25	166764	37563591
216	4	1297	12	227.065	105	23841.825
216	5	1297	12	490.89	2239	1099102.71
216	6	1297	12	466.19	495578	231033507.82
217	0	1303	12	250.43	586921	146982626.03

Query 2:

```
/*Dice by Oats and Manitoba*/
WITH crop_id AS (
    SELECT * FROM public."Crop" WHERE "Common Name" = 'Oats'
), region_id AS (
    SELECT * FROM public."Region" WHERE "Name_EN" = 'Manitoba'
)
SELECT F.*
FROM public."FactTable" F, crop_id, region_id
WHERE F."Crop_ID" = crop_id."Crop_ID" AND F."Region_ID" = region_id."Region_ID";
```

Output for Query 2:

Date_ID	Crop_ID	Climate_ID	Region_ID	Price	Volume	Total_Value
0	2	3	8	94.51	81042	7659279.42
1	2	9	8	87.91	38423	3377765.93
2	2	15	8	97.93	40098	3926797.14
3	2	21	8	92.9	24479	2274099.1
4	2	27	8	94.24	24453	2304450.72
5	2	33	8	98.65	23851	2352901.15
6	2	39	8	120.29	51479	6192408.91
7	2	45	8	124.47	78777	9805373.19

Combining OLAP operations [4 queries]

Query 1:

```
-- Compare total yearly productions of Wheat
select
  sum(F."Volume") as Total_Volume,
  D."Year"
from
  "FactTable" as F,
  "Crop" as C,
  "Date" as D
where
  F."Crop_ID" = C."Crop_ID"
  and F."Date_ID" = D."Date_ID"
  and C."Common Name" = 'Wheat' -- slice by the Crop 'wheat'
group by
  D."Year"; -- roll-up by year
```

Output for Query 1:

total_volume	Year
689193.601908066	2001
17426	2002
26203	2003
69823	2004
108281	2005
108938	2006
110823	2007
103581	2008
104055	2009
122374	2010

Query 2:

```
-- Compare total productions in each quarter of Oats across different regions
select
  sum(F."Volume") as Total_Volume,
  R."GeographicRegion",
  D."Quarter"
from
  "FactTable" as F,
  "Crop" as C,
  "Region" as R,
  "Date" as D
where
  F."Crop_ID" = C."Crop_ID"
  and F."Region_ID" = R."Region_ID"
  and F."Date_ID" = D."Date_ID"
  and C."Common Name" = 'Oats' -- slice by the Crop 'oats'
group by
  R."GeographicRegion", -- roll-up by region
  D."Quarter"; -- roll-up by quarter
```

Output for Query 2:

total_volume	GeographicRegior	Quarter
236049	"British Columb	1
192681	"British Columb	2
211766.5	"British Columb	3
263348	"British Columb	4
828371	"Canada"	1
802279	"Canada"	2
1573321	"Canada"	3
988530.5	"Canada"	4
12430633	"Prairies"	1
9780085	"Prairies"	2
15768459	"Prairies"	3
14308467	"Prairies"	4

Query 3:

```
-- Find average volume of Barley produced in different Temperature bands
select
  avg(F."Volume") as Average_Volume,
  Cl."Mean_Temp_Binned",Cl."Mean_Temp_Binned_Encoded"
from
  "FactTable" as F,
  "Crop" as C,
  "Climate" as Cl
where
  F."Crop_ID" = C."Crop_ID"
  and F."Climate_ID" = Cl."Climate_ID"
  and C."Common Name" = 'Barley' -- slice by Barley
  Cl."Mean_Temp_Binned",Cl."Mean_Temp_Binned_Encoded" -- roll-up by mean temperatures
order by
  Cl."Mean_Temp_Binned_Encoded" DESC
```

Output for Query 3:

average_volume	Mean_Temp_Binned	Mean_Temp_Binned_Encoded
29003.6707317073	"(19.347, 21.861]"	40
41600.9875	"(18.601, 19.347]"	39
41542.5365853659	"(18.077, 18.601]"	38
51974.8	"(17.664, 18.077]"	37
40642.925	"(17.182, 17.664]"	36
43409.4390243902	"(16.653, 17.182]"	35
39213.8	"(16.191, 16.653]"	34
64873.3625	"(15.514, 16.191]"	33
49983.7073170732	"(15.052, 15.514]"	32

Query 4:

```
-- Find total Rye value in Saskatchewan in each band of ‘Total Precip’
select
  sum(F."Total_Value") as Sum_Total_Values,
  Cl."Total_Precip_Binned",Cl."Total_Precip_Binned_Encoded"
from
  "FactTable" as F,
  "Crop" as C,
  "Region" as R,
  "Date" as D,
  "Climate" as Cl
where
  F."Crop_ID" = C."Crop_ID"
  and F."Region_ID" = R."Region_ID"
  and F."Date_ID" = D."Date_ID"
  and F."Climate_ID" = Cl."Climate_ID"
  and C."Common Name" = 'Rye' -- dice by Crop and Province
  and R."Name_EN" = 'Saskatchewan'
group by
  Cl."Total_Precip_Binned",Cl."Total_Precip_Binned_Encoded" -- rollup by precipitation
bands
order by
  Cl."Total_Precip_Binned_Encoded" DESC
```

Output for Query 4:

sum_total_values	Total_Precip_Binned	Total_Precip_Binned_Encoded
1189880.52	"(113.279, 128.736]"	38
530261.05	"(100.693, 106.354]"	36
4004269.85	"(91.588, 95.206]"	34
2573347.73	"(87.882, 91.588]"	33
1062729.25	"(84.872, 87.882]"	32
2665858.95	"(80.661, 84.872]"	31
787540	"(75.39, 78.082]"	29
5584132.05	"(73.277, 75.39]"	28
5044175.99	"(70.504, 73.277]"	27

Part 2. Explorative Operations

Iceberg Queries

Query: Top five largest Total Value in a month across all recorded

```
-- Top five largest Total Value in a month across all recorded
select
  F."Total_Value",
  C."Common Name",
  R."Name_EN",
  D."Date"
from
  "FactTable" as F,
  "Crop" as C,
  "Region" as R,
  "Date" as D
where
  F."Crop_ID" = C."Crop_ID"
  and F."Region_ID" = R."Region_ID"
  and F."Date_ID" = D."Date_ID"
```

```
order by
  F."Total_Value" DESC
limit 5
```

Output for Query:

Total_Value	Common Name	Name_EN	Date
995129106.89	"Canola"	"Saskatchewan"	"2023-03-01"
985891409.6	"Canola"	"Saskatchewan"	"2022-11-01"
890134312.32	"Canola"	"Saskatchewan"	"2023-01-01"
880016082.09	"Canola"	"Saskatchewan"	"2021-09-01"
868109676.1	"Canola"	"Saskatchewan"	"2022-09-01"

Windowing Queries

Query: Filter by province of Ontario, partition by type of grain, compute each grain's average value (across all datapoints) and create rank for months within a particular grain

```
-- Filter by province of Ontario, partition by type of grain, compute each grain's
average value (across all datapoints) and create rank for months within a particular
grain

SELECT
  C."Common Name",
  D."Date",
  "Total_Value",
  AVG("Total_Value") OVER (PARTITION BY C."Crop_ID") AS "Average_Value",
  RANK() OVER (PARTITION BY C."Crop_ID" ORDER BY "Total_Value" DESC) AS "Rank"
FROM
  "FactTable" as F,
  "Region" as R,
  "Date" as D,
  "Crop" as C
WHERE
 F."Crop_ID" = C."Crop_ID"
  and F."Region_ID" = R."Region_ID"
  and F."Date_ID" = D."Date_ID"
```

```
  AND R."Name_EN" = 'Ontario'
```

Output for Query:

Common Name	Date	Total_Value	Average_Value	Rank
"Oats"	"2015-09-01"	7652755.5	1021763.26876812	1
"Oats"	"2022-08-01"	7230924.84	1021763.26876812	2
"Oats"	"2022-09-01"	5938922.64	1021763.26876812	3
"Oats"	"2020-09-01"	5037798.36	1021763.26876812	4
"Oats"	"2023-09-01"	4967545.44	1021763.26876812	5
"Oats"	"2021-09-01"	4518879.36	1021763.26876812	6
"Oats"	"2019-09-01"	4326731.28	1021763.26876812	7

Using the Window clause

Query: Moving window on Volume in Ontario by crop

```
select
   C."Common Name",
   R."Name_EN",
   D."Date",
   AVG(F."Total_Value") OVER W AS movavg_value
from
  "FactTable" as F,
  "Crop" as C,
  "Region" as R,
  "Date" as D
where
  F."Crop_ID" = C."Crop_ID"
  and F."Region_ID" = R."Region_ID"
  and F."Date_ID" = D."Date_ID"
  and R."Name_EN" = 'Ontario'
window w as (partition by C."Common Name"
  order by D."Date"
  range between interval '1' month preceding
  and interval '1' month following)
```

Output of Query:

Common Name	Name_EN	Date	movavg_value
"Barley"	"Ontario"	"2001-01-01"	872347.875
"Barley"	"Ontario"	"2001-02-01"	873346.6
"Barley"	"Ontario"	"2001-03-01"	894177.15
"Barley"	"Ontario"	"2001-04-01"	917663.65
"Barley"	"Ontario"	"2001-05-01"	926531.45
"Barley"	"Ontario"	"2001-06-01"	897228.2
"Barley"	"Ontario"	"2001-07-01"	854469.6
"Barley"	"Ontario"	"2001-08-01"	849838.15

Part B: BI Dashboard and Information Visualization

The dashboard can be found in the *dashboard* folder in the zipped folder and on the GitHub repository.

The following dashboard was created on Power BI to explore the data and visualize trends:

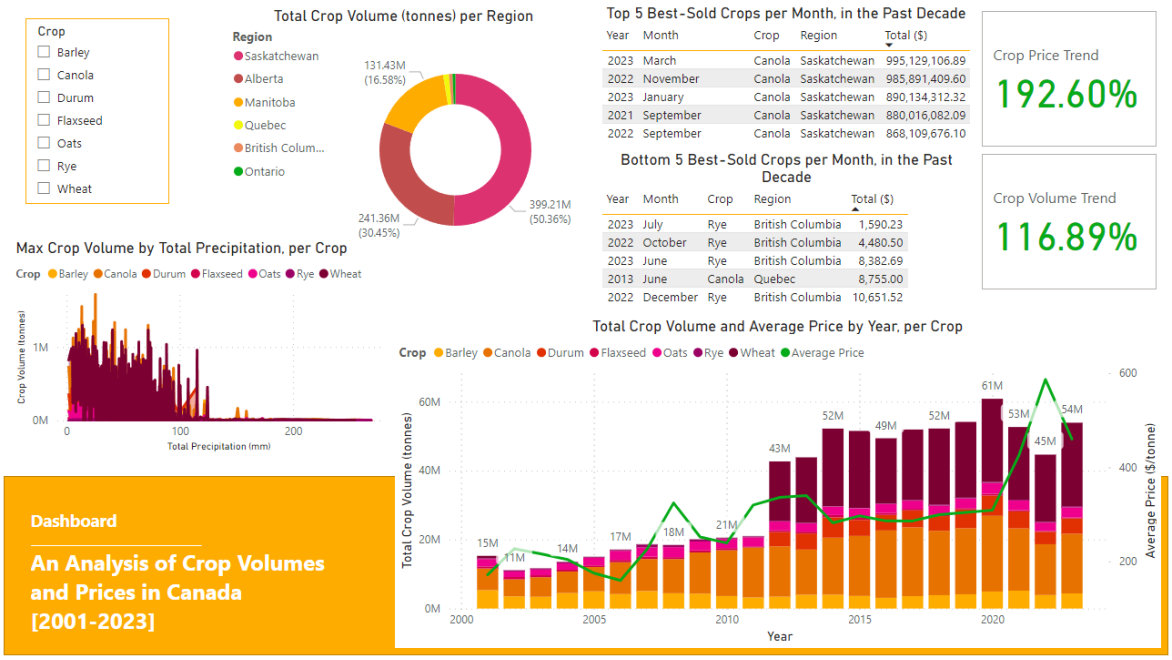

Figure 1: Dashboard for the Analysis of Crop Volumes and Prices in Canada in Power BI.

To visualize the general trend for price and volume for all crops and for all regions in Canada over the entire period of 2001-2023, a line and stacked column chart was used (refer to the "Total Crop Volume and Average Price by Year, per Crop" chart). From it, we can extract the general trend for price and volume metrics by using the following formula:

Trend = (Latest value – Earliest value) / Earliest value

These trend metrics are recorded on the top right of the dashboard and conditionally formatted to easily differentiate positive (marked in green) and negative (marked in red) trends.

The same general trend graph could also be used to transverse the date concept hierarchy – by clicking on a value on the price line, the graph will zoom into the quarter view for that year, and with another click on some quarter, the graph will zoom into the monthly view for the quarter (Figure 2). Likewise, you can zoom out into a less complex view though the upwards arrow on the side of the graph. As such, you can drill-down and roll-up using the date hierarchy in this graph.

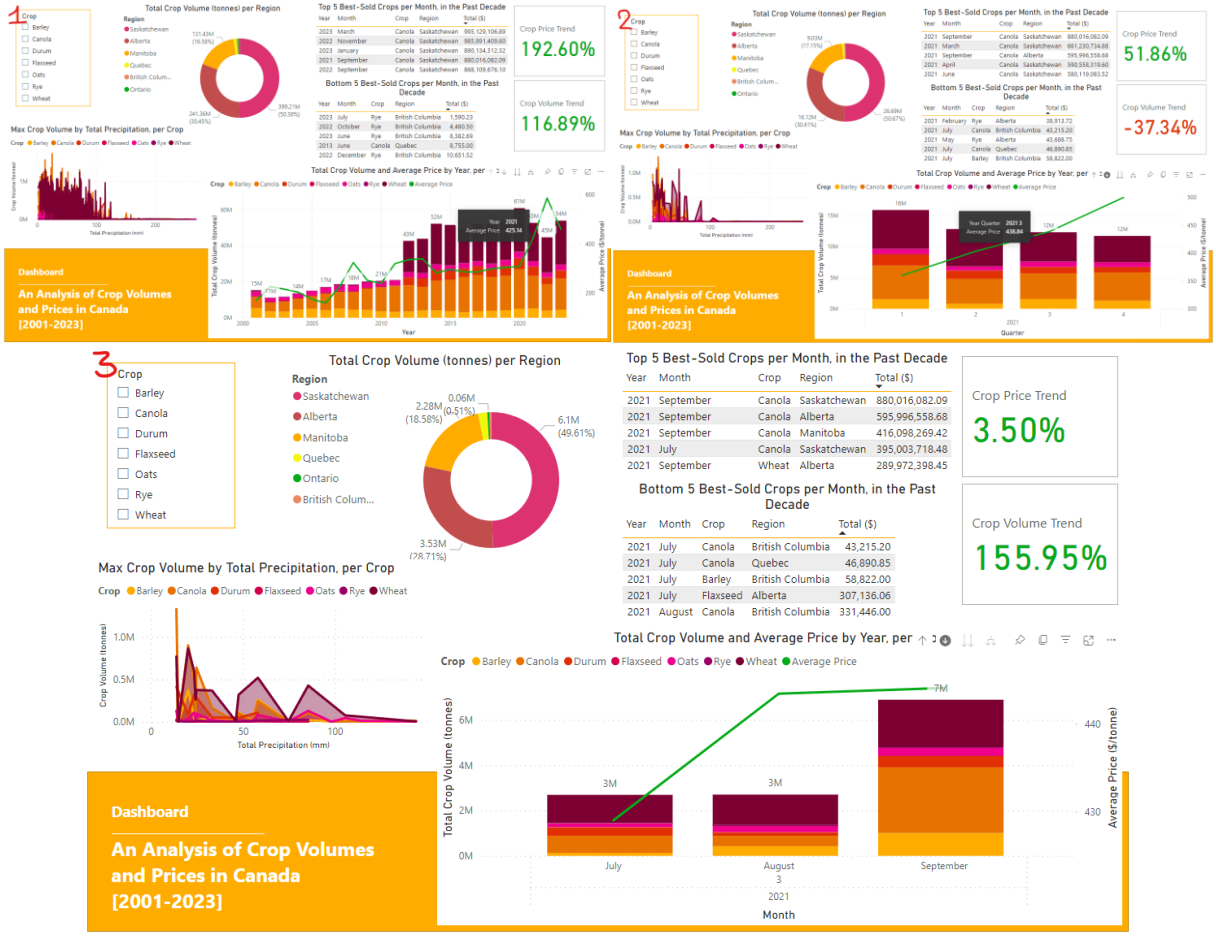

Figure 2: Drill-down example for the date hierarchy, from year (1) to quarter (2) to month (3).

To slice the data and update all the visualizations accordingly, the list of crops was provided on the top left. By selecting a crop, all the visualizations are automatically updated to display data for only that crop (Figure 3).

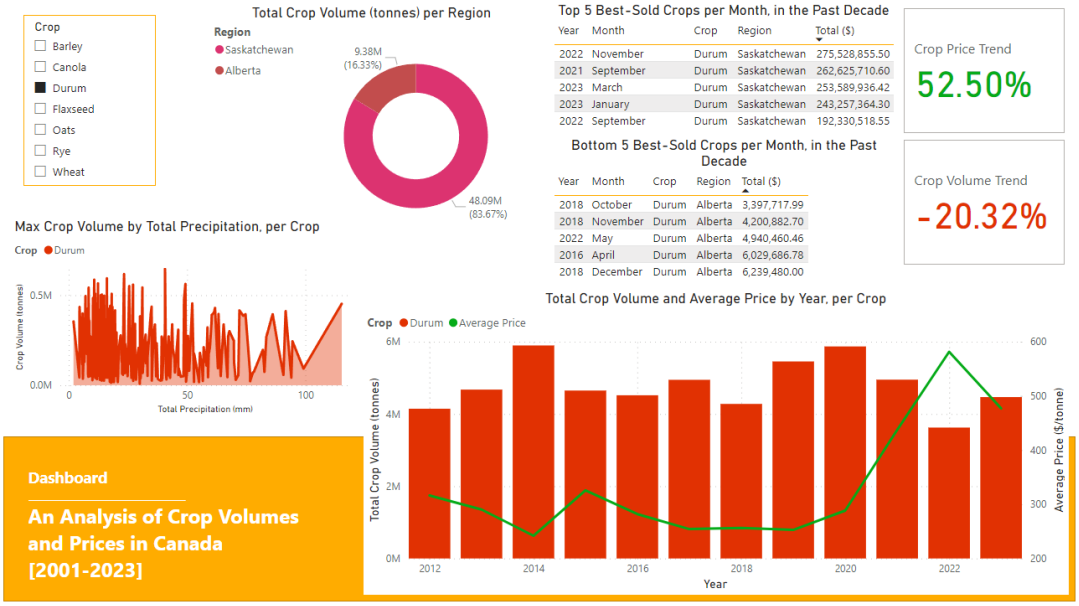

Figure 3: Slicing example with the Crop set to "Durum"

To further dice the data, we can select the crop and one of the regions in the legend of the nearby donut chart that displays the volume of crops produced by region (Figure 4). Upon selecting a crop and region, all the visualizations will once again automatically update to display the related data only.

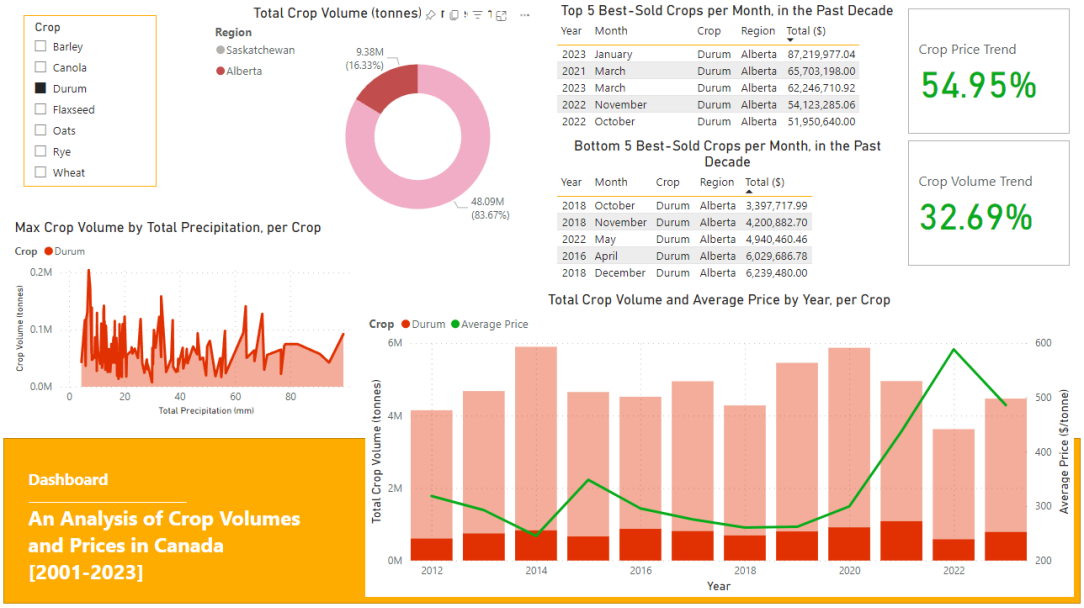

Figure 4: Dicing example with the Crop set to "Durum" and the Region set to "Alberta"

The two tables located in the top center of the dashboard are used to display the top 5 and bottom 5 total production values (the product of price and volume), and were added to clearly display the related crops, regions and dates to the most- and least-sold.

Lastly, the final 2 graphs on the dashboard are the Donut chart and the Area chart. The Donut chart, located in the top center of the dashboard, displays the total crop volume per region, explaining what region accounts for most of that crop's production. The Area chart, located on the left of the dashboard, displays the max crop volume achieved for a certain precipitation. This graph, through historical data, can be used to assess the attainable crop volume, based on weather patterns.