Warning: package 'RSQLite' was built under R version 4.3.3 #connect to the database conn <- dbConnect(RSQLite::SQLite()," FinalDB.sqlite")</pre> dbExecute(conn, "DROP TABLE IF EXISTS CROP_DATA") [1] 0 dbExecute(conn, "DROP TABLE IF EXISTS DAILY_FX") [1] 0 dbExecute(conn, "DROP TABLE IF EXISTS FARM_DATA") [1] 0 dbExecute(conn, "DROP TABLE IF EXISTS MONTHLY_FX") [1] 0 #Load the data into the database #CD_ID, YEAR, CROP_TYPE, GEO, SEEDED_AREA, HARVESTED_AREA, PRODUCTION, AVG_YIELD df1 <- dbExecute(conn,</pre> "CREATE TABLE CROP_DATA (CD_ID INTEGER NOT NULL, YEAR DATE NOT NULL, CROP_TYPE VARCHAR(20) NOT NULL, GEO VARCHAR(20) NOT NULL, SEEDED_AREA INTEGER NOT NULL, HARVESTED_AREA INTEGER NOT NULL, PRODUCTION INTEGER NOT NULL, AVG_YIELD INTEGER NOT NULL, PRIMARY KEY (CD_ID) if (df1 == -1){ cat ("An error has occurred.\n") msg <- odbcGetErrMsg(conn)</pre> print (msg) } else { cat ("Table was created successfully.\n") Table was created successfully. #DFX_ID,DATE,FXUSDCAD df2 <- dbExecute(conn, "CREATE TABLE DAILY_FX (</pre> DFX_ID INTEGER NOT NULL, DATE DATE NOT NULL, FXUSDCAD FLOAT(6), PRIMARY KEY (DFX_ID) if (df2 == -1){ cat ("An error has occurred.\n") msg <- odbcGetErrMsg(conn)</pre> print (msg) } else { cat ("Table was created successfully.\n") Table was created successfully. #FARM_PRICES #CD_ID,DATE,CROP_TYPE,GEO,PRICE_PRERMT df3 <- dbExecute(conn,</pre> "CREATE TABLE FARM_DATA (CD_ID INTEGER NOT NULL, DATE DATE NOT NULL, CROP_TYPE VARCHAR(20) NOT NULL, GEO VARCHAR(20) NOT NULL, PRICE_PRERMT INTEGER NOT NULL, PRIMARY KEY (CD_ID) if (df2 == -1){ cat ("An error has occurred.\n") msg <- odbcGetErrMsg(conn)</pre> print (msg) } else { cat ("Table was created successfully.\n") Table was created successfully. #MONTHLY_FX #DFX_ID,DATE,FXUSDCAD df4 <- dbExecute(conn, "CREATE TABLE MONTHLY_FX (</pre> DFX_ID INTEGER NOT NULL, DATE DATE NOT NULL, FXUSDCAD FLOAT(6), PRIMARY KEY (DFX_ID) if (df2 == -1){ cat ("An error has occurred.\n") msg <- odbcGetErrMsg(conn)</pre> print (msg) } else { cat ("Table was created successfully.\n") Table was created successfully. Crop_Data <- read.csv('https://cf-courses-data.s3.us.cloud-object-storage.appdomain.clou</pre> Farm_p_prices<- read.csv('https://cf-courses-data.s3.us.cloud-object-storage.appdomain.c</pre> D_FX_Data<- read.csv('https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud</pre> M_FX_Data<- read.csv('https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud head(Crop_Data) YEAR CROP_TYPE CD_ID GEO SEEDED_AREA HARVESTED_AREA PRODUCTION 1372000 Barley 1372000 2504000 0 1965-12-31 Alberta 2476800 2476800 4752900 1 1965–12–31 Barley Canada 2 1965–12–31 708000 1415000 Barley Saskatchewan 708000 3 1965–12–31 297400 215500 297400 Canola Alberta 4 1965–12–31 580700 580700 512600 Canola Canada 5 1965–12–31 Canola Saskatchewan 242700 224600 224600 AVG_YIELD 1825 1920 2000 725 885 1080 head(Farm_p_prices) DATE CROP_TYPE GEO PRICE_PRERMT CD_ID 127.39 0 1985-01-01 Barley Alberta 121.38 1 1985-01-01 Barley Saskatchewan 342.00 2 1985-01-01 Canola Alberta Canola Saskatchewan 339.82 3 1985-01-01 100.77 4 1985-01-01 Alberta Rye 5 1985-01-01 109.75 Rye Saskatchewan head(D_FX_Data) DFX_ID DATE FXUSDCAD 0 2017-01-03 1.3435 1 2017-01-04 1.3315 2 2017-01-05 1.3244 3 2017-01-06 1.3214 4 2017-01-09 1.3240 5 2017-01-10 1.3213 head(M_FX_Data) DFX_ID DATE FXUSDCAD 0 2017-01-01 1.319276 1 2017-02-01 1.310726 2 2017-03-01 1.338643 3 2017-04-01 1.344021 4 2017-05-01 1.360705 5 2017-06-01 1.329805 #creating a table with the csv dbWriteTable(conn, "CROP_DATA", Crop_Data, overwrite=TRUE, row.names = FALSE) dbWriteTable(conn, "MONTHLY_FX", M_FX_Data, overwrite=TRUE, row.names = FALSE) dbWriteTable(conn, "FARM_DATA", Farm_p_prices, overwrite=TRUE, row.names = FALSE) dbWriteTable(conn, "DAILY_FX", D_FX_Data, overwrite=TRUE, row.names = FALSE) # List the tables in the database dbListTables(conn) [1] "CROP_DATA" "DAILY_FX" "FARM_DATA" "MONTHLY_FX" # Retrieve the first 5 rows from each table dbGetQuery(conn, 'SELECT * from CROP_DATA LIMIT 5') YEAR CROP_TYPE GEO SEEDED_AREA HARVESTED_AREA PRODUCTION $\mathsf{CD}\mathsf{_ID}$ 0 1965-12-31 Barley 1372000 1372000 2504000 Alberta 1 1965–12–31 2476800 2476800 4752900 Barley Canada 1415000 2 1965–12–31 708000 708000 Barley Saskatchewan 3 1965–12–31 297400 297400 215500 Canola Alberta 4 1965–12–31 Canola 580700 512600 Canada 580700 AVG_YIELD 1825 1920 2000 725 885 dbGetQuery(conn, "SELECT * from MONTHLY_FX LIMIT 5") DFX_ID DATE FXUSDCAD 0 2017-01-01 1.319276 1 2017-02-01 1.310726 2 2017-03-01 1.338643 3 2017-04-01 1.344021 4 2017-05-01 1.360705

dbGetQuery(conn, "SELECT * from DAILY_FX LIMIT 5")

dbGetQuery(conn, "SELECT * from FARM_DATA LIMIT 5")

Barley

Canola

Which geographies are included in the farm prices dataset?

Farm price just have 2 locations, Alberta and Saskatchewan.

How many hectares of Rye were harvested in Canada in 1968?

In 1968 in Canada it was harvested a total area of 274100 hectares.

274100

from FARM_DATA WHERE CROP_TYPE = 'Rye' LIMIT 6")

Barley was produced for both provinces on the data, Alberta and Saskatchewan.

from FARM_DATA WHERE CROP_TYPE = 'Barley' ")

dbGetQuery(conn, "select DISTINCT(GEO) AS PROVINCES, CROP_TYPE AS PRODUCE

The first date priced for a produce was January first of 1985 and the last was on December first of 2020.

dbGetQuery(conn, "select MIN(DATE) AS FIRST_DATE, MAX(DATE) AS LAST_DATE FROM FARM_DATA"

The only crop that reached the higest price greather than or equal to \$350 per metric tonne was Canola.

dbGetQuery(conn, "select DISTINCT(CROP_TYPE) AS PRODUCE, PRICE_PRERMT FROM FARM_DATA WHE

dbGetQuery(conn, "select CROP_TYPE AS PRODUCE, PRICE_PRERMT FROM FARM_DATA WHERE PRICE_P

Rank the crop types harvested in Saskatchewan in the year 2000 by their average yield. Which crop

dbGetQuery(conn, "select DISTINCT(CROP_TYPE) PRODUCE, GEO AS LOCATION, AVG_YIELD, strfti

Rank the crops and geographies by their average yield (KG per hectare) since the year 2000. Which crop

dbGetQuery(conn, "select CROP_TYPE, GEO, AVG_YIELD, strftime('%Y',YEAR) AS YEAR FROM CRO

Use a subquery to determine how much wheat was harvested in Canada in the most recent year of the

The recent year of the data is 2020. It was harvested about 10017800 hectares of wheat in Canada.

FROM CROP_DATA WHERE CROP_TYPE ="Wheat" AND GEO = "Canada" ')

10017800

Use an implicit inner join to calculate the monthly price per metric tonne of Canola grown in

Saskatchewan in both Canadian and US dollars. Display the most recent 6 months of the data.

dbGetQuery(conn, "SELECT CROP_TYPE, strftime('%Y',YEAR) AS YEAR, GEO, DATE, FXUSDCAD, PR

WHERE GEO = 'Saskatchewan' AND CROP_TYPE = 'Canola' AND strftime('%Y',YEAR) =

850241

850241

850241

850241

850241

850241

DATE FXUSDCAD MONTHLY_PROD

The best crop harvested in Saskatchewan in 2000 was Barley with the avg yield of 2800.

Which crops have ever reached a farm price greater than or equal to \$350 per metric tonne?

Query and display the first 6 rows of the farm prices table for Rye.

dbGetQuery(conn, "select PRICE_PRERMT, CROP_TYPE

Rye

Rye

Rye

Rye

Rye

Rye

Find the first and last dates for the farm prices data

Rye

GEO PRICE_PRERMT

127.39

121.38

342.00

339.82

100.77

Alberta

Alberta

Alberta

dbGetQuery(conn, "select count(cd_id) as Total_records from FARM_DATA")

dbGetQuery(conn, "select distinct(geo) as geografic_locations from FARM_DATA")

dbGetQuery(conn, "select strftime('%Y',YEAR) AS YEAR, SEEDED_AREA, HARVESTED_AREA from C

Barley Saskatchewan

Canola Saskatchewan

DATE FXUSDCAD

DATE CROP_TYPE

How many records are in the farm prices dataset?

The total records in the farm are 2,678

0 2017-01-03 1.3435

1 2017-01-04 1.3315

2 2017-01-05 1.3244

3 2017-01-06 1.3214

4 2017-01-09 1.3240

DFX_ID

CD_ID

0 1985-01-01

1 1985-01-01

2 1985-01-01

3 1985-01-01

4 1985-01-01

Total_records

2678

geografic_locations

1 1968

Alberta

YEAR SEEDED_AREA HARVESTED_AREA

274100

PRICE_PRERMT CROP_TYPE

100.77

109.75

95.05

103.46

96.77

106.38

Which provinces grew Barley?

PROVINCES PRODUCE

2 Saskatchewan Barley

FIRST_DATE LAST_DATE

PRODUCE PRICE_PRERMT

350.04

363.84

364.36

354.58

359.99

350.47

354.05

354.82

357.37

368.18

127.39

342.00

100.77

273.83

LOCATION AVG_YIELD YEAR

and province had the highest average yield since the year 2000?

GEO AVG_YIELD YEAR

Barley produced in Alberta had the highest avg yield in the year of 2013.

4100 2013

4100 2016

3980 2020

3900 2013

3900 2016

3900 2016

3900 2017

3890 2019

3820 2020

3810 2019

GEO, MAX(YEAR), HARVESTED_AREA

GEO MAX(YEAR) HARVESTED_AREA

GE0

Canola 2020 Saskatchewan 2020-12-01 1.280771

Canola 2020 Saskatchewan 2020-11-01 1.306820

Canola 2020 Saskatchewan 2020-10-01 1.321471

Canola 2020 Saskatchewan 2020-09-01 1.322810

Canola 2020 Saskatchewan 2020-08-01 1.322205

Canola 2020 Saskatchewan 2020-07-01 1.349850

2800 2000

2200 2000

2100 2000

1400 2000

Canola

Canola

Canola

Canola

Canola

Canola

Canola

Canola

Canola

PRODUCE PRICE_PRERMT

10 Canola

1 Barley

2 Canola

Rye

Wheat

performed best?

PRODUCE

1 Barley Saskatchewan

Wheat Saskatchewan

4 Canola Saskatchewan

CROP_TYPE

data

CROP_TYPE

CROP_TYPE YEAR

Price_Metric_Tone

663850.7

650618.3

643404.8

642753.9

643047.8

629878.1

Barley Alberta

Barley Alberta

Barley Alberta

Wheat Alberta

Barley Canada

Wheat Alberta

Barley Alberta

Barley Alberta

Barley Canada

Barley Canada

dbGetQuery(conn, 'SELECT CROP_TYPE,

Wheat Canada 2020-12-31

FROM CROP_DATA

JOIN MONTHLY_FX

Rye Saskatchewan

1 1985-01-01 2020-12-01

Alberta Barley

Saskatchewan

Peer Reviewed Assignment

REAL DATA SQL AND ANALYSIS

#install.packages("https://cran.r-project.org/src/contrib/Archive/RSQLite/RSQLite_0.10.0

AUTHOR

salisa almeida

library("RSQLite")