Zoe Erpelding

Final Project

CPSC324 Spring 2024

Implementation

(d). Implement your project and submit all related code that you developed.

Preview the heart disease data:

SELECT \*

FROM `cpsc324-final-project.project\_datasets.heart`

A screenshot of a computer

Description automatically generated

A white sheet with black text

Description automatically generatedA screenshot of a computer

Description automatically generated

Preview the insurance data:

SELECT \*

FROM `cpsc324-final-project.project\_datasets.insurance`

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Heart disease and insurance combined on state:

SELECT Indicator, State, GeographicLevel, LocationDesc, Value, Quartile\_Number, Year, Data\_Value as HeartDiseaseRate\_per\_Hundred\_Thousand, Stratification1 as Gender, Stratification2 as Race\_Ethnicity, Y\_lat, X\_lon

FROM `cpsc324-final-project.project\_datasets.insurance` as i JOIN `cpsc324-final-project.project\_datasets.heart` as h ON i.State = h.LocationDesc

WHERE Class = 'Cardiovascular Diseases' AND Topic = 'Heart Disease Mortality' AND State != 'United States'

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Insurance data formatting (splitting the city from the state):

SELECT Year, SPLIT(GeoName, ',') [safe\_offset(0)] AS city, SPLIT(GeoName, ',') [safe\_offset(1)] as state, Percapita\_unemployment\_insurance\_compensation, Wages\_and\_salaries\_avg, Wages\_and\_salaries, Unemployment\_insurance, Percapita\_net\_earnings, Percapita\_personal\_income, Population, Earnings\_per\_job\_avg, Personal\_income, Total\_employment/Population as employment\_rate

FROM `cpsc324-final-project.project\_datasets.income\_iffy`

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Average heart disease mortality rate per state:

SELECT LocationDesc, AVG(Data\_Value)/100000 AS heart\_disease\_rate

FROM `cpsc324-final-project.project\_datasets.heart`

WHERE GeographicLevel = 'State'

GROUP BY LocationDesc

ORDER BY heart\_disease\_rate DESC

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Clean Heart Dataset

SELECT Year, LocationAbbr, LocationDesc, Data\_Value AS Mortality\_Rate, Stratification1 AS Gender, Stratification2 AS Race, Y\_lat, X\_lon

FROM `cpsc324-final-project.project\_datasets.heart`

WHERE GeographicLevel = 'County' AND Topic = 'Heart Disease Mortality'AND StratificationCategory1 = 'Gender' AND StratificationCategory2 = 'Race/Ethnicity'

A screenshot of a computer

Description automatically generated

Looker Studio: Heart

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

<https://lookerstudio.google.com/reporting/05fd86e2-2042-45bc-8074-c7946617d851>

A screenshot of a medical insurance application

Description automatically generated

<https://lookerstudio.google.com/reporting/0d4779aa-f726-43a6-9824-10731371bcf2>

BQML model:

CREATE OR REPLACE MODEL `project\_datasets.model`

OPTIONS(model\_type='linear\_reg', input\_label\_cols=['HeartDiseaseRate\_per\_Hundred\_Thousand']) AS

  SELECT \*

  FROM `project\_datasets.insurance\_heart`

  WHERE HeartDiseaseRate\_per\_Hundred\_Thousand IS NOT NULL

  LIMIT 2000

SELECT \*

FROM ml.evaluate(MODEL `project\_datasets.model`,(

  SELECT \*

  FROM `project\_datasets.insurance\_heart`

  WHERE HeartDiseaseRate\_per\_Hundred\_Thousand IS NOT NULL

  LIMIT 2000

)

)

A screenshot of a computer

Description automatically generated

CREATE OR REPLACE TABLE `project\_datasets.predictions`

AS (

  SELECT \*

  FROM ml.predict(MODEL `project\_datasets.model`, (

    SELECT \*

    FROM `project\_datasets.insurance\_heart`

    LIMIT 100

  ))

)

A screenshot of a computer

Description automatically generated

Try it with no limit:

CREATE OR REPLACE MODEL `project\_datasets.model`

OPTIONS(model\_type='linear\_reg', input\_label\_cols=['HeartDiseaseRate\_per\_Hundred\_Thousand']) AS

  SELECT \*

  FROM `project\_datasets.insurance\_heart`

  WHERE HeartDiseaseRate\_per\_Hundred\_Thousand IS NOT NULL

SELECT \*

FROM ml.evaluate(MODEL `project\_datasets.model`,(

  SELECT \*

  FROM `project\_datasets.insurance\_heart`

  WHERE HeartDiseaseRate\_per\_Hundred\_Thousand IS NOT NULL

)

)

A screenshot of a computer

Description automatically generated

CREATE OR REPLACE TABLE `project\_datasets.predictions`

AS (

  SELECT predicted\_HeartDiseaseRate\_per\_Hundred\_Thousand, HeartDiseaseRate\_per\_Hundred\_Thousand, ABS(predicted\_HeartDiseaseRate\_per\_Hundred\_Thousand-HeartDiseaseRate\_per\_Hundred\_Thousand) AS difference, State, Gender, Race\_Ethnicity

  FROM ml.predict(MODEL `project\_datasets.model`, (

    SELECT \*

    FROM `project\_datasets.insurance\_heart`

    LIMIT 100

  ))

  ORDER BY difference DESC

)

A screenshot of a computer

Description automatically generated

CREATE OR REPLACE TABLE `project\_datasets.predictions`

AS (

  SELECT predicted\_HeartDiseaseRate\_per\_Hundred\_Thousand, HeartDiseaseRate\_per\_Hundred\_Thousand, ABS(predicted\_HeartDiseaseRate\_per\_Hundred\_Thousand-HeartDiseaseRate\_per\_Hundred\_Thousand) AS difference, State, Gender, Race\_Ethnicity

  FROM ml.predict(MODEL `project\_datasets.model`, (

    SELECT \*

    FROM `project\_datasets.insurance\_heart`

    LIMIT 100

  ))

  ORDER BY difference ASC

)

A screenshot of a computer

Description automatically generated