**SQL QUERY FOR DATA ANALYSIS**

First, convert the data file into CSV file from and Next, complete the following steps in BigQuery console to upload the data.

**Step 1:** Open BigQuery console and click on the project you want to upload the data to.

**Step 2:** In the Explorer on the left, click the Actions icon (three vertical dots) next to your project name and select **Create dataset**.

**Step 3:** Name "electrical\_vehicles" will be used for the dataset. Enter **electrical\_vehicles** for the Dataset ID.

**Step 4:** Click **CREATE DATASET** (blue button) to add the dataset to your project.

**Step 5:** In the Explorer on the left, click to expand your project and then click the **electrical\_vehicles** dataset you just created.

**Step 6:** Click the Actions icon (three vertical dots) next to electrical\_vehicles and select **Open**.

**Step 7:** Click the blue + icon at the top right to open the Create table window.

**Step 8:** Under Source, for the Create table from selection, choose where the data will be coming from.

* Select **Upload**.
* Click **Browse** to select the Movie Data CSV file you downloaded.
* Choose **CSV** from the file format drop-down.

**Step 9:** Under Destination, for Table name enter **electrical\_vehicles**.

**Step 10:** For Schema, click the Auto detect check box.

**Step 11**: Click **Create table** (blue button). You will now see the **electrical\_vehicles** table under your **electrical\_vehicles** dataset in your project.

**Step 12:** Click **electrical\_vehicles** and then select the **Preview** tab. Confirm that you see the data shown below.

**To preview :**

SELECT  \*

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

**To change from categorical data to numerical data :**

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET gender =  CASE

WHEN gender = 'Male' THEN 1

WHEN gender = 'Female' THEN 2

ELSE 3 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET age =  CASE

WHEN age = 'Below 25' THEN 1

WHEN age = '25 - 35' THEN 2

WHEN age = '35 - 45' THEN 3

ELSE 4 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET occupation = CASE

WHEN occupation = 'Student' THEN 1 WHEN occupation = 'Government employee' THEN 2 WHEN occupation = 'Private employee' THEN 3 WHEN occupation = 'Self employee' THEN 4

ELSE 5 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET  = education\_level\_ CASE

WHEN education\_level\_ = 'High School' THEN 1

WHEN education\_level\_ = 'Bachelors Degree' THEN 2 WHEN education\_level\_ = 'Masters Degree' THEN 3

ELSE 4 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = interested\_in\_ele\_tw\_ec CASE

WHEN interested\_in\_ele\_tw\_ec = 'Highly' THEN 1

WHEN interested\_in\_ele\_tw\_ec = 'A little' THEN 2

ELSE 3END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = view\_on\_ev CASE

WHEN view\_on\_ev = 'Well Known' THEN 1

ELSE 2 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = driven\_an\_ev CASE

WHEN driven\_an\_ev = 'Yes' THEN 1

ELSE 2 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = owned\_an\_ev CASE

WHEN owned\_an\_ev = 'Yes' THEN 1

ELSE 2 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = prefer\_ev\_or\_fv CASE

WHEN prefer\_ev\_or\_fv = 'Electrical Vehicle' THEN 1

ELSE 2 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = known\_about\_ban CASE

WHEN known\_about\_ban = 'Yes' THEN 1

ELSE 2 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = are\_you\_concerned\_about\_this\_ban CASE

WHEN are\_you\_concerned\_about\_this\_ban = 'Strongly Agree' THEN 1

WHEN are\_you\_concerned\_about\_this\_ban = 'Agree' THEN 2

WHEN are\_you\_concerned\_about\_this\_ban = 'Disagree' THEN 3

ELSE 4 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = happy\_with\_ev CASE

WHEN happy\_with\_ev = 'Strongly Agree' THEN 1 WHEN happy\_with\_ev = 'Agree' THEN 2

WHEN happy\_with\_ev = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = ev\_are\_much\_quitter CASE

WHEN ev\_are\_much\_quitter = 'Strongly Agree' THEN 1

WHEN hev\_are\_much\_quitter = 'Agree' THEN 2 WHEN ev\_are\_much\_quitter = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = excellent\_acceleration CASE

WHEN excellent\_acceleration = 'Strongly Agree' THEN 1 WHEN excellent\_acceleration = 'Agree' THEN 2 WHEN excellent\_acceleration = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = environmentally\_friendly\_ CASE

WHEN environmentally\_friendly\_ = 'Strongly Agree' THEN 1 WHEN environmentally\_friendly\_ = 'Agree' THEN 2 WHEN environmentally\_friendly\_ = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = cost\_of\_ev CASE

WHEN cost\_of\_ev = 'Strongly Agree' THEN 1

WHEN cost\_of\_ev = 'Agree' THEN 2

WHEN cost\_of\_ev = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = ev\_have\_much\_better\_way CASE

WHEN ev\_have\_much\_better\_way = 'Strongly Agree' THEN 1

WHEN ev\_have\_much\_better\_way = 'Agree' THEN 2 WHEN ev\_have\_much\_better\_way = 'Disagree' THEN 3

ELSE 4 END;

UPDATE  `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = global\_warming\_ CASE WHEN global\_warming\_ = 'Strongly Agree' THEN 1

WHEN global\_warming\_ = 'Agree' THEN 2

WHEN global\_warming\_ = 'Disagree' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = factor\_encourage\_ CASE

WHEN factor\_encourage\_ = 'Price' THEN 1 WHEN factor\_encourage\_ = 'Positive environment' THEN 2

WHEN factor\_encourage\_ = 'Newtrend' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = factors\_discourage CASE

WHEN factors\_discourage = 'limited range' THEN 1

WHEN factors\_discourage = 'long recharging time' THEN 2 WHEN factors\_discourage = 'Lack of trust to new technologies' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = charging\_points CASE

WHEN charging\_points = 'Each 50 KM' THEN 1

WHEN charging\_points = 'Each 100 KM' THEN 2 WHEN charging\_points = 'Each 150 KM' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = tesla CASE

WHEN tesla = 'Deep knowledge of features' THEN 1

WHEN tesla = 'Superficial awareness' THEN 2

WHEN tesla = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = bmw\_iseries CASE

WHEN bmw\_iseries = 'Deep knowledge of features' THEN 1

WHEN bmw\_iseries = 'Superficial awareness' THEN 2 WHEN bmw\_iseries = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = renault\_zoe CASE  WHEN renault\_zoe = 'Deep knowledge of features' THEN 1

WHEN renault\_zoe = 'Superficial awareness' THEN 2 WHEN renault\_zoe = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = chevrolet\_volt CASE

WHEN chevrolet\_volt = 'Deep knowledge of features' THEN 1

WHEN chevrolet\_volt = 'Superficial awareness' THEN 2

WHEN chevrolet\_volt = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = cadillac\_poolside CASE

WHEN cadillac\_poolside = 'Deep knowledge of features' THEN 1

WHEN cadillac\_poolside = 'Superficial awareness' THEN 2

WHEN cadillac\_poolside = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = toyota\_prius CASE

WHEN toyota\_prius = 'Deep knowledge of features' THEN 1

WHEN toyota\_prius = 'Superficial awareness' THEN 2

WHEN toyota\_prius = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = hyundai\_kona CASE

WHEN hyundai\_kona = 'Deep knowledge of features' THEN 1

WHEN hyundai\_kona = 'Superficial awareness' THEN 2 WHEN hyundai\_kona = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = tata\_nexon CASE

WHEN tata\_nexon = 'Deep knowledge of features' THEN 1

WHEN tata\_nexon = 'Superficial awareness' THEN 2

WHEN tata\_nexon = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = Mahindra CASE

WHEN mahindra = 'Deep knowledge of features' THEN 1

WHEN mahindra = 'Superficial awareness' THEN 2 WHEN mahindra = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = mg CASE

WHEN mg = 'Deep knowledge of features' THEN 1

WHEN mg = 'Superficial awareness' THEN 2

WHEN mg = 'Sounds familiar' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = knowledge\_about\_ev CASE

WHEN knowledge\_about\_ev = 'Newspapers' THEN 1 WHEN knowledge\_about\_ev = 'Magazine' THEN 2

WHEN knowledge\_about\_ev = 'Television' THEN 3

ELSE 4 END;

UPDATE `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicles`

SET = buying\_an\_ev CASE

WHEN buying\_an\_ev = 'Will definitely buy one' THEN 1

WHEN buying\_an\_ev = 'An likely to buy one' THEN 2

WHEN buying\_an\_ev = 'Am considering buying one but need convincing' THEN 3

WHEN buying\_an\_ev = 'Definitely wont buy' THEN 4

ELSE 5 END;

**To find Descriptive Statistics for Gender :**

**For Mean :**

SELECT AVG(gender) AS average\_value

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`.

**For Median** :

SELECT gender AS median\_value

FROM (SELECT gender, ROW\_NUMBER() OVER (ORDER BY gender) AS row\_num,

         COUNT(\*) OVER () AS total\_rows

  FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) sub

WHERE row\_num = (total\_rows + 1) / 2;

**For Mode :**

SELECT gender AS mode\_value, COUNT(\*) AS frequency

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY gender HAVING COUNT(\*) = (SELECT MAX(freq)

  FROM (SELECT gender, COUNT(\*) AS freq

    FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

    GROUP BY gender) sub);

**For Sum :**

SELECT SUM(gender) AS total\_sum

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**For Standard Error :**

SELECT SQRT(VARIANCE(gender) / COUNT(\*) ) AS standard\_error

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**For Standard Deviation :**

SELECT SQRT(SUM((gender - avg\_value) \* (gender - avg\_value)) / (COUNT(\*) - 1)) AS standard\_deviation

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

CROSS JOIN (SELECT AVG(gender) AS avg\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) AS subquery;

**For Sample Variance :**

SELECT

  (SUM((gender - mean\_value) \* (gender - mean\_value)) / (COUNT(\*) - 1)) AS sample\_variance

FROM

  (SELECT AVG(gender) AS mean\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) subquery, `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**For count :**

SELECT COUNT(\*) AS total\_records

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle` ;

**For frequency and percentage :**

SELECT gender, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY gender;

**To find Descriptive Statistics for Age :**

**To find mean:**

SELECT AVG(age) AS average\_value

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

**To find Median:**

SELECT age AS median\_value

FROM (

  SELECT age, ROW\_NUMBER() OVER (ORDER BY age) AS row\_num,

         COUNT(\*) OVER () AS total\_rows

  FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

) sub

WHERE row\_num = (total\_rows + 1) / 2;

**To find Mode :**

SELECT age AS mode\_value, COUNT(\*) AS frequency

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY age

HAVING COUNT(\*) = (

  SELECT MAX(freq)

  FROM (

    SELECT age, COUNT(\*) AS freq

    FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

    GROUP BY age

  ) sub

);

**To find Sum:**

SELECT SUM(age) AS total\_sum

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To Standard Error :**

SELECT SQRT(VARIANCE(age) / COUNT(\*) ) AS standard\_error

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To Standard Deviation:**

SELECT SQRT(SUM((age - avg\_value) \* (age - avg\_value)) / (COUNT(\*) - 1)) AS standard\_deviation

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

CROSS JOIN (SELECT AVG(age) AS avg\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) AS subquery;

**To Sample variance:**

SELECT

  (SUM((age - mean\_value) \* (age - mean\_value)) / (COUNT(\*) - 1)) AS sample\_variance

FROM

  (SELECT AVG(age) AS mean\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) subquery, `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To count:**

SELECT COUNT(\*) AS total\_records

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle` ;

**To frequency and percentage :**

SELECT age, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY age;

**To find Descriptive statistics for occupation:**

**To find mean:**

SELECT AVG(occupation) AS average\_value

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

**To find median:**

SELECT occupation AS median\_value

FROM (

  SELECT occupation, ROW\_NUMBER() OVER (ORDER BY occupation) AS row\_num,

         COUNT(\*) OVER () AS total\_rows

  FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

) sub

WHERE row\_num = (total\_rows + 1) / 2;

**To find mode:**

SELECT occupation AS mode\_value, COUNT(\*) AS frequency

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY occupation

HAVING COUNT(\*) = (

  SELECT MAX(freq)

  FROM (

    SELECT occupation, COUNT(\*) AS freq

    FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

    GROUP BY occupation

  ) sub

);

**To find sum:**

SELECT SUM(occupation) AS total\_sum

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To standard error:**

SELECT SQRT(VARIANCE(occupation) / COUNT(\*) ) AS standard\_error

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To find standard deviation:**

SELECT SQRT(SUM((occupation - avg\_value) \* (occupation - avg\_value)) / (COUNT(\*) - 1)) AS standard\_deviation

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

CROSS JOIN (SELECT AVG(occupation) AS avg\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) AS subquery;

**To sample variance:**

SELECT

  (SUM((occupation - mean\_value) \* (occupation - mean\_value)) / (COUNT(\*) - 1)) AS sample\_variance

FROM

  (SELECT AVG(occupation) AS mean\_value FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`) subquery, `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`;

**To count:**

  SELECT COUNT(\*) AS total\_records

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle` ;

**To frequency and percentage:**

SELECT occupation, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY occupation;

**Global warming:**

SELECT global\_warming\_, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*)

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY global\_warming\_;

**Knowledge:**

SELECT knowledge\_about\_ev, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*)

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY knowledge\_about\_ev;

**Charging**

SELECT charging\_points, COUNT(\*) AS frequency,

       (COUNT(\*) \* 100.0 / (SELECT COUNT(\*)

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`)) AS percentage

FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

GROUP BY charging\_points;

**Correlation:**

WITH statistics AS (

  SELECT

    AVG(factor\_encourage\_) AS mean\_x,

    AVG(factors\_discourage) AS mean\_y,

    STDDEV(factor\_encourage\_) AS stddev\_x,

    STDDEV(factors\_discourage) AS stddev\_y,

    COVAR\_POP(factor\_encourage\_, factors\_discourage) AS covariance

  FROM `electrical-vehicles-386123.electrical\_vehicles.electrical\_vehicle`

)

SELECT

  covariance / (stddev\_x \* stddev\_y) AS pearson\_correlation

FROM statistics;