SQL Bike Store Analysis  
Pasted from PgAdmin. The .sql file is available from the GitHub repository.

--First I created the Database

CREATE DATABASE Bike Store;

DROP TABLE IF EXISTS orders;

--Then, I created the tables of the sales from the Excel Database.

CREATE TABLE customers(

customer\_id SMALLINT PRIMARY KEY,

first\_name VARCHAR(100),

last\_name VARCHAR(100),

email VARCHAR(100),

street VARCHAR(100),

city VARCHAR(100),

state VARCHAR(100),

zip\_code INT

);

CREATE TABLE stores(

store\_id SMALLINT PRIMARY KEY,

store\_name VARCHAR(100),

phone VARCHAR(100),

email VARCHAR(100),

street VARCHAR(100),

city VARCHAR(100),

state VARCHAR(100),

zip\_code INT

);

CREATE TABLE staffs(

staff\_id SMALLINT PRIMARY KEY,

first\_name VARCHAR(100),

last\_name VARCHAR(100),

email VARCHAR(100),

phone VARCHAR(100),

active VARCHAR(100),

store\_id INT,

FOREIGN KEY (store\_id) REFERENCES stores (store\_id)

);

CREATE TABLE orders(

order\_id SMALLINT PRIMARY KEY,

customer\_id INT,

order\_status VARCHAR(100),

order\_date DATE,

shipped\_date DATE,

store\_id INT,

staff\_id INT,

FOREIGN KEY (customer\_id) REFERENCES customers (customer\_id),

FOREIGN KEY (store\_id) REFERENCES stores (store\_id),

FOREIGN KEY (staff\_id) REFERENCES staffs (staff\_id)

);

--Then I uploaded the information from the Bike Store Sales.

--Customers

COPY customers (customer\_id, first\_name, last\_name,

email, street, city, state, zip\_code)

FROM 'C:\Users\Public\Public Doc4SQL\dataset\customers.csv' DELIMITER ','

CSV Header ;

--Stores

COPY stores (store\_id, store\_name, phone,

email, street, city, state, zip\_code)

FROM 'C:\Users\Public\Public Doc4SQL\dataset\stores.csv' DELIMITER ','

CSV Header ;

--Staffs

COPY staffs (staff\_id, first\_name, last\_name,

email, phone, active, store\_id)

FROM 'C:\Users\Public\Public Doc4SQL\dataset\staffs.csv' DELIMITER ','

CSV Header ;

--Orders

COPY orders (order\_id, customer\_id, order\_status, order\_date,

shipped\_date, store\_id, staff\_id)

FROM 'C:\Users\Public\Public Doc4SQL\dataset\orders.csv' DELIMITER ','

CSV Header ;

--Then I verified everything individually using the SELECT command

SELECT \* FROM customers;

SELECT \* FROM stores;

SELECT \* FROM staffs;

SELECT \* FROM orders;

--After the creation of these table I conducted various analysis

--I conducted a customer order analysis by store, to allow us to see which customers made what orders at what store.

SELECT c.customer\_id, c.first\_name, s.store\_name,

o.order\_id, o.order\_date, o.shipped\_date

FROM customers c

INNER JOIN orders o ON c.customer\_id = o.customer\_id

INNER JOIN stores s ON o.store\_id = s.store\_id;

--Then I showed the amount of orders by customer, the max being 3.

SELECT c.customer\_id, c.first\_name, c.last\_name, COUNT(o.order\_id) AS total\_orders

FROM customers c

LEFT JOIN orders o ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_id, c.first\_name, c.last\_name

ORDER BY total\_orders DESC;

--Here, I identified the sales per store and state which these shops were in.

SELECT s.store\_id, s.store\_name, s.state, COUNT(o.order\_id) AS total\_orders

FROM stores s

LEFT JOIN orders o ON s.store\_id = o.store\_id

GROUP BY s.store\_id, s.store\_name, s.state

ORDER BY total\_orders DESC;

-- From this we can see that Baldwin bikes in NY is the best performing shop based on number of orders.

-- Staff Analysis

--First I wanted to know the amount of staff per store

SELECT s.store\_id, st.store\_name, COUNT(\*) AS staff\_count

FROM staffs s

JOIN stores st ON s.store\_id = st.store\_id

GROUP BY s.store\_id, st.store\_name

ORDER BY staff\_count DESC;

--The most staff members are in the Santa Cruiz location.

--TO analyse the staff further, and see if we are overstaffed, I displayed the sales by staff member showing their location.

SELECT s.staff\_id, s.first\_name, s.last\_name, st.city, COUNT(DISTINCT o.customer\_id) AS total\_customers\_served

FROM staffs s

LEFT JOIN orders o ON s.staff\_id = o.staff\_id

LEFT JOIN stores st ON s.store\_id = st.store\_id

GROUP BY s.staff\_id, s.first\_name, s.last\_name, st.city

ORDER BY total\_customers\_served DESC;

-- From this we can see the top 2 staff members are from Baldwin, with a significant difference from the other stores.

--This is concerning due to difference in sales rates in Santa cruiz.

--They should consider why this is, as the total customers served by two staff members from santa monica is 0.

--Below displays the top employees in each store.

WITH RankedEmployees AS (

SELECT s.store\_id, s.staff\_id, s.first\_name, s.last\_name, COUNT(o.order\_id) AS total\_orders,

ROW\_NUMBER() OVER (PARTITION BY s.store\_id ORDER BY COUNT(o.order\_id) DESC) AS rank

FROM staffs s

LEFT JOIN orders o ON s.staff\_id = o.staff\_id

GROUP BY s.store\_id, s.staff\_id, s.first\_name, s.last\_name

)

SELECT re.store\_id, re.staff\_id, re.first\_name, re.last\_name, re.total\_orders, st.store\_name

FROM RankedEmployees re

JOIN stores st ON re.store\_id = st.store\_id

WHERE re.rank = 1 ORDER BY re.total\_orders DESC ;

--Distribution analysis by store, here I showed the average delivery processing time by store

SELECT s.store\_id, s.store\_name, AVG(o.shipped\_date - o.order\_date) AS avg\_processing\_time

FROM stores s

INNER JOIN orders o ON s.store\_id = o.store\_id

GROUP BY s.store\_id, s.store\_name;

-- This showed me the best average time was Baldwin bikes.

--My analysis provides a comprehensive view of customer behaviour, store performance, and staff efficiency. Allowing for informed decision-making to optimise business operations and improve customer satisfaction.