SP Rocket Dataset

Business Data Management Final Project Report

Submitted in partial fulfilment of the requirements for the award of the degree of

Masters in

Information Technology and Analytics

By

Ashita Shetty
ashita.shetty@rutgers.edu
Manan Desai
md1810@scarletmail.rutgers.edu
Gauri Raghuram
gr514@scarletmail.rutgers.edu
Krishna Niveditha Pallem
kp1160@scarletmail.rutgers.edu
Sanskritee Rajpal
sr1926@scarletmail.rutgers.edu

Master's in Information Technology and Analytics

Rutgers Business School, Newark

December 2023

TASK AND ACOUNTABILITY CONTRACT

Work Breakdown	Ashita Shetty	Manan Desai	Gauri Raghuram	Krishna Niveditha	Sanskritee Rajpal
Search and create datasets	✓		✓	✓	
ER Diagram				✓	
Create SQL code and import datasets to database	✓			✓	
Dataset cleaning with SQL code	✓	✓			✓
Design your own SQL question and write SQL queries	✓		✓		
Presentation and report		✓			✓
Visualizations			✓	✓	

ER DIAGRAM

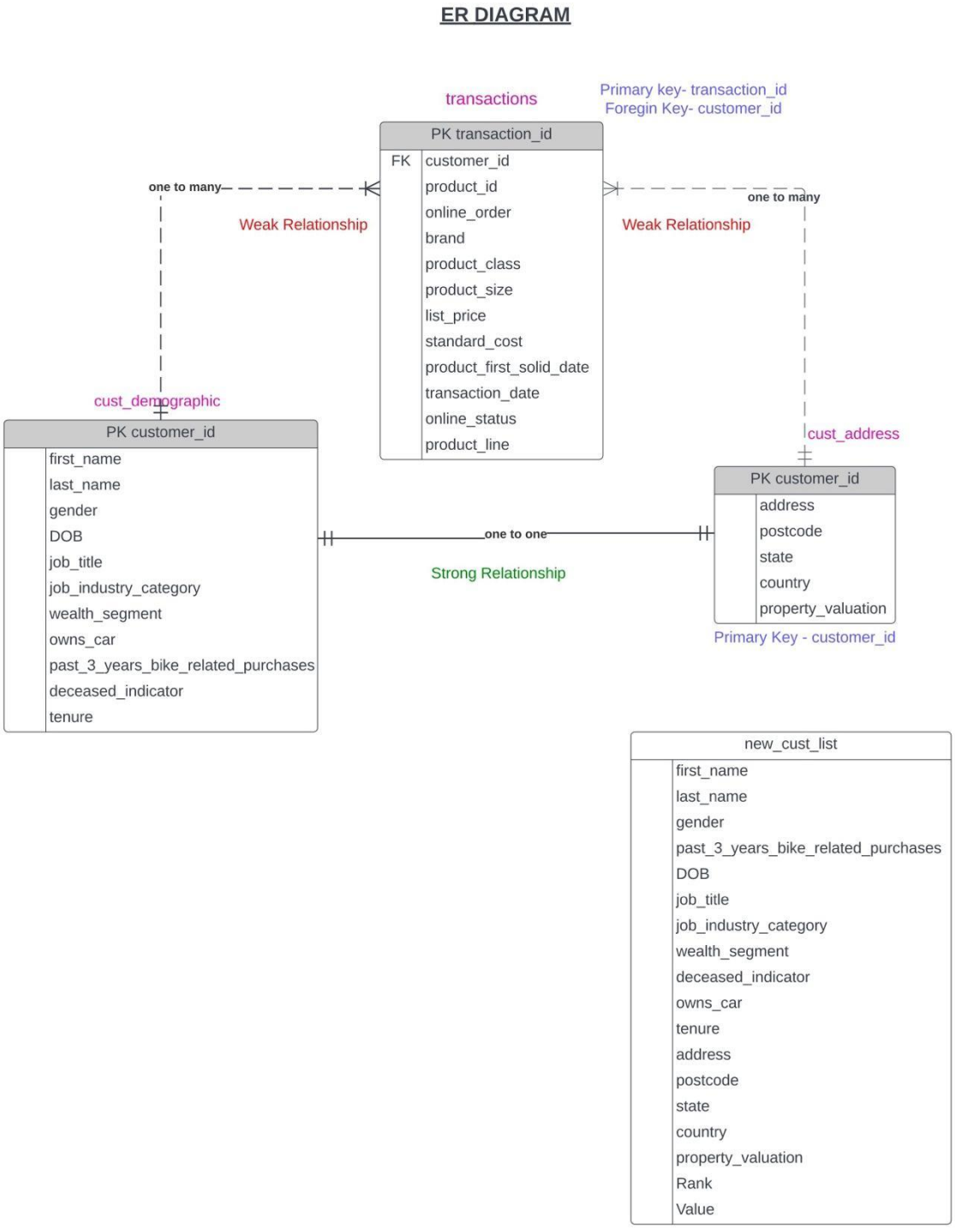

ER DIAGRAM
transactions
Primary key- transaction_id
Foregin Key- customer_id
PK transaction_id
FK customer_id
product_id
online_order
brand
product_class
product_size
list_price
standard_cost
product_first_solid_date
transaction_date
online_status
product_line
one to many
Weak Relationship
one to many
Weak Relationship
cust_demographic
PK customer_id
first_name
last_name
gender
DOB
job_title
job_industry_category
wealth_segment
owns_car
past_3_years_bike_related_purchases
deceased_indicator
tenure
one to one
Strong Relationship
cust_address
PK customer_id
address
postcode
state
country
property_valuation
Primary Key - customer_id
new_cust_list
first_name
last_name
gender
past_3_years_bike_related_purchases
DOB
job_title
job_industry_category
wealth_segment
deceased_indicator
owns_car
tenure
address
postcode
state
country
property_valuation
Rank
Value

TABLE CREATION

Code:
Transactions Table

```
/* Transactions */

CREATE TABLE transactions
(
    transaction_id  INT,
    product_id  INT,
    customer_id INT,
    transaction_date    VARCHAR(512),
    online_order    VARCHAR(512),
    order_status    VARCHAR(512),
    brand   VARCHAR(512),
    product_line    VARCHAR(512),
    product_class   VARCHAR(512),
    product_size    VARCHAR(512),
    list_price  FLOAT,
    standard_cost   VARCHAR(512),
    product_first_sold_date VARCHAR(512)
);


INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('1', '2', '2950',
'2/25/17', 'FALSE', 'Approved', 'Solex', 'Standard', 'medium', 'medium', '71.49', '$53.62', '41245');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('2', '3', '3120',
'5/21/17', 'TRUE', 'Approved', 'Trek Bicycles', 'Standard', 'medium', 'large', '2091.47', '$388.92', '41701');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('3', '37', '402',
'10/16/17', 'FALSE', 'Approved', 'OHM Cycles', 'Standard', 'low', 'medium', '1793.43', '$248.82', '36361');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('4', '88', '3135',
'8/31/17', 'FALSE', 'Approved', 'Norco Bicycles', 'Standard', 'medium', 'medium', '1198.46', '$381.10', '36145');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('5', '78', '787',
'10/1/17', 'TRUE', 'Approved', 'Giant Bicycles', 'Standard', 'medium', 'large', '1765.3', '$709.48', '42226');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('6', '25', '2339',
'3/8/17', 'TRUE', 'Approved', 'Giant Bicycles', 'Road', 'medium', 'medium', '1538.99', '$829.65', '39031');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('7', '22', '1542',
'4/21/17', 'TRUE', 'Approved', 'WeareA2B', 'Standard', 'medium', 'medium', '60.34', '$45.26', '34165');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('8', '15', '2459',
'7/15/17', 'FALSE', 'Approved', 'WeareA2B', 'Standard', 'medium', 'medium', '1292.84', '$13.44', '39915');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('9', '67', '1305',
'8/10/17', 'FALSE', 'Approved', 'Solex', 'Standard', 'medium', 'large', '1071.23', '$380.74', '33455');
INSERT INTO transactions (transaction_id, product_id, customer_id, transaction_date, online_order, order_status, brand,
product_line, product_class, product_size, list_price, standard_cost, product_first_sold_date) VALUES ('10', '12', '3262',
'8/30/17', 'TRUE', 'Approved', 'WeareA2B', 'Standard', 'medium', 'medium', '1231.15', '$161.60', '38216');

CREATE TABLE cust_address
(
customer_id           INT,
address    VARCHAR(512),
postcode  INT,
state         VARCHAR(512),
country    VARCHAR(512),
property_valuation INT
);
```

```
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('1', '060 Morning Avenue', '2016', 'New South Wales', 'Australia', '10');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('2', '6 Meadow Vale Court', '2153', 'New South Wales', 'Australia', '10');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('4', '0 Holy Cross Court', '4211', 'QLD', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('5', '17979 Del Mar Point', '2448', 'New South Wales', 'Australia', '4');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('6', '9 Oakridge Court', '3216', 'VIC', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('7', '4 Delaware Trail', '2210', 'New South Wales', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('8', '49 Londonderry Lane', '2650', 'New South Wales', 'Australia', '4');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('9', '97736 7th Trail', '2023', 'New South Wales', 'Australia', '12');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('11', '93405 Ludington Park', '3044', 'VIC', 'Australia', '8');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('12', '44339 Golden Leaf Alley', '4557', 'QLD', 'Australia', '4');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('13', '2 Sutherland Street', '3799', 'VIC', 'Australia', '6');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('14', '9 Mcbride Trail', '2760', 'New South Wales', 'Australia', '8');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('15', '9861 New Castle Avenue', '2428', 'New South Wales', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('16', '52 Moland Street', '3331', 'VIC', 'Australia', '4');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('17', '82391 Kensington Lane', '3058', 'VIC', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('18', '092 2nd Alley', '2135', 'New South Wales', 'Australia', '12');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('19', '59 Spaight Circle', '2233', 'New South Wales', 'Australia', '9');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('20', '032 Bartelt Crossing', '2444', 'New South Wales', 'Australia', '8');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('21', '18 Jenna Center', '2650', 'New South Wales', 'Australia', '7');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('24', '3 Cordelia Plaza', '2153', 'New South Wales', 'Australia', '10');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('25', '28 5th Center', '4413', 'QLD', 'Australia', '3');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('26', '52 Carey Alley', '4740', 'QLD', 'Australia', '5');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('27', '96 Texas Plaza', '3218', 'VIC', 'Australia', '7');
INSERT INTO cust_address (customer_id, address, postcode, state, country, property_valuation) VALUES ('28', '48 Eagan Avenue', '4868', 'QLD', 'Australia', '4');


CREATE TABLE new_cust_list
(
first_name              VARCHAR(512),
last_name               VARCHAR(512),
gender      VARCHAR(512),
past_3_years_bike_related_purchases     INT,
DOB         VARCHAR(512),
job_title   VARCHAR(512),
job_industry_category           VARCHAR(512),
wealth_segment      VARCHAR(512),
deceased_indicator VARCHAR(512),
owns_car VARCHAR(512),
tenure      INT,
address     VARCHAR(512),
postcode  INT,
state       VARCHAR(512),
country     VARCHAR(512),
```

```
property_valuation  VARCHAR(512),
Rank        INT,
Value       VARCHAR(512)
);


INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Chickie', 'Brister', 'Male', '86', '1957-07-12', 'General Manager',
'Manufacturing', 'Mass Customer', 'N', 'Yes', '14', '45 Shopko Center', '4500', 'QLD', 'Australia', '6', '1', '1.71875');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Morly', 'Genery', 'Male', '69', '1970-03-22', 'Structural Engineer', 'Property',
'Mass Customer', 'N', 'No', '16', '14 Mccormick Park', '2113', 'NSW', 'Australia', '11', '1', '1.71875');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Ardelis', 'Forrester', 'Female', '10', '1974-08-28', 'Senior Cost Accountant',
'Financial Services', 'Affluent Customer', 'N', 'No', '10', '5 Colorado Crossing', '3505', 'VIC', 'Australia', '5', '1', '1.71875');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Lucine', 'Stutt', 'Female', '64', '1979-01-28', 'Account Representative III',
'Manufacturing', 'Affluent Customer', 'N', 'Yes', '5', '207 Annamark Plaza', '4814', 'QLD', 'Australia', '1', '4', '1.703125');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Melinda', 'Hadlee', 'Female', '34', '1965-09-21', 'Financial Analyst', 'Financial
Services', 'Affluent Customer', 'N', 'No', '19', '115 Montana Place', '2093', 'NSW', 'Australia', '9', '4', '1.703125');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Druci', 'Brandli', 'Female', '39', '1951-04-29', 'Assistant Media Planner',
'Entertainment', 'High Net Worth', 'N', 'Yes', '22', '89105 Pearson Terrace', '4075', 'QLD', 'Australia', '7', '6', '1.671875');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Rutledge', 'Hallt', 'Male', '23', '1976-10-06', 'Compensation Analyst', 'Financial
Services', 'Mass Customer', 'N', 'No', '8', '7 Nevada Crossing', '2620', 'NSW', 'Australia', '7', '6', '1.671875');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Nancie', 'Vian', 'Female', '74', '1972-12-27', 'Human Resources Assistant II',
'Retail', 'Mass Customer', 'N', 'Yes', '10', '85 Carioca Point', '4814', 'QLD', 'Australia', '5', '8', '1.65625');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Duff', 'Karlowicz', 'Male', '50', '1972-04-28', 'Speech Pathologist',
'Manufacturing', 'Mass Customer', 'N', 'Yes', '5', '717 West Drive', '2200', 'NSW', 'Australia', '10', '8', '1.65625');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Barthel', 'Docket', 'Male', '72', '1985-08-02', 'Accounting Assistant IV', 'IT',
'Mass Customer', 'N', 'Yes', '17', '80 Scofield Junction', '4151', 'QLD', 'Australia', '5', '10', '1.640625');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Rockwell', 'Matson', 'Male', '94', '1995-01-01', 'Programmer Analyst I', 'Retail',
'High Net Worth', 'N', 'No', '3', '3682 Crowley Point', '4573', 'QLD', 'Australia', '6', '10', '1.640625');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Wheeler', 'Winward', 'Male', '48', '1999-08-30', 'Environmental Specialist',
'Manufacturing', 'Mass Customer', 'N', 'No', '10', '3 Golden Leaf Point', '3216', 'VIC', 'Australia', '8', '12', '1.625');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Olag', '', 'Male', '60', '1990-05-13', 'Human Resources Manager',
'Telecommunications', 'Mass Customer', 'N', 'No', '9', '0484 North Avenue', '2032', 'NSW', 'Australia', '11', '13', '1.609375');
INSERT INTO new_cust_list (first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title,
job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure, address, postcode, state, country,
property_valuation, Rank, Value) VALUES ('Melba', 'Spellacy', 'Female', '38', '1976-12-09', 'VP Marketing', 'Health', 'Mass
Customer', 'N', 'No', '4', '0591 Anzinger Circle', '2232', 'NSW', 'Australia', '10', '14', '1.59375');


CREATE TABLE cust_demographic
(
customer_id             INT,
```

```
first_name              VARCHAR(512),
last_name               VARCHAR(512),
gender      VARCHAR(512),
past_3_years_bike_related_purchases     INT,
DOB         VARCHAR(512),
job_title   VARCHAR(512),
job_industry_category           VARCHAR(512),
wealth_segment      VARCHAR(512),
deceased_indicator VARCHAR(512),
owns_car VARCHAR(512),
tenure      VARCHAR(512)
);

INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('1', 'Laraine',
'Medendorp', 'F', '93', '1953-10-12', 'Executive Secretary', 'Health', 'Mass Customer', 'N', 'Yes', '11');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('2', 'Eli',
'Bockman', 'Male', '81', '1980-12-16', 'Administrative Officer', 'Financial Services', 'Mass Customer', 'N', 'Yes', '16');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('3', 'Arlin',
'Dearle', 'Male', '61', '1954-01-20', 'Recruiting Manager', 'Property', 'Mass Customer', 'N', 'Yes', '15');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('4', 'Talbot', '',
'Male', '33', '1961-10-03', '', 'IT', 'Mass Customer', 'N', 'No', '7');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('5', 'Sheila-
kathryn', 'Calton', 'Female', '56', '1977-05-13', 'Senior Editor', 'n/a', 'Affluent Customer', 'N', 'Yes', '8');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('6', 'Curr',
'Duckhouse', 'Male', '35', '1966-09-16', '', 'Retail', 'High Net Worth', 'N', 'Yes', '13');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('7', 'Fina',
'Merali', 'Female', '6', '1976-02-23', '', 'Financial Services', 'Affluent Customer', 'N', 'Yes', '11');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('8', 'Rod', 'Inder',
'Male', '31', '1962-03-30', 'Media Manager I', 'n/a', 'Mass Customer', 'N', 'No', '7');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('9', 'Mala', 'Lind',
'Female', '97', '1973-03-10', 'Business Systems Development Analyst', 'Argiculture', 'Affluent Customer', 'N', 'Yes', '8');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('10', 'Fiorenze',
'Birdall', 'Female', '49', '1988-10-11', 'Senior Quality Engineer', 'Financial Services', 'Mass Customer', 'N', 'Yes', '20');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('11', 'Uriah',
'Bisatt', 'Male', '99', '1954-04-30', '', 'Property', 'Mass Customer', 'N', 'No', '9');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('12', 'Sawyere',
'Flattman', 'Male', '58', '1994-07-21', 'Nuclear Power Engineer', 'Manufacturing', 'Mass Customer', 'N', 'No', '8');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('13', 'Gabriele',
'Norcross', 'Male', '38', '1955-02-15', 'Developer I', 'Financial Services', 'High Net Worth', 'N', 'Yes', '8');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('14', 'Rayshell',
'Kitteman', 'Female', '85', '1983-03-25', 'Account Executive', 'Financial Services', 'Affluent Customer', 'N', 'No', '6');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('15', 'Erroll',
'Radage', 'Male', '91', '2000-07-13', 'Junior Executive', 'Manufacturing', 'Mass Customer', 'N', 'No', '1');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('16', 'Harlin',
'Parr', 'Male', '38', '1977-02-27', 'Media Manager IV', 'n/a', 'Mass Customer', 'N', 'Yes', '18');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('17', 'Heath',
'Faraday', 'Male', '57', '1962-03-19', 'Sales Associate', 'n/a', 'Affluent Customer', 'N', 'Yes', '15');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases,
DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('18', 'Marjie',
'Neasham', 'Female', '79', '1967-07-06', 'Professor', 'n/a', 'Affluent Customer', 'N', 'No', '11');
```

```
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('19', 'Sorcha', 'Keyson', 'Female', '76', '2001-04-15', 'Geological Engineer', 'Manufacturing', 'High Net Worth', 'N', 'No', '1');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('20', 'Basile', 'Firth', 'Male', '72', '1980-08-13', 'Project Manager', 'Manufacturing', 'Mass Customer', 'N', 'No', '11');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('21', 'Mile', 'Cammocke', 'Male', '74', '1980-09-20', 'Safety Technician I', 'Manufacturing', 'Affluent Customer', 'N', 'Yes', '16');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('22', 'Deeanne', 'Durtnell', 'Female', '79', '1962-12-10', '', 'IT', 'Mass Customer', 'N', 'No', '11');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('23', 'Olav', 'Polak', 'Male', '43', '1995-02-10', '', 'n/a', 'High Net Worth', 'N', 'Yes', '1');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('24', 'Kim', 'Skpsey', 'Female', '55', '1977-12-03', 'Research Assistant I', 'Argiculture', 'Mass Customer', 'N', 'Yes', '15');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('25', 'Geoff', 'Assaf', 'Male', '72', '1976-12-02', 'Accounting Assistant III', 'Financial Services', 'Mass Customer', 'N', 'Yes', '21');
INSERT INTO cust_demographic (customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_segment, deceased_indicator, owns_car, tenure) VALUES ('26', 'Trixi', 'Ginnelly', 'Female', '12', '1978-06-10', 'Editor', 'Financial Services', 'Mass Customer', 'N', 'Yes', '16');
```

DATA CLEANING

```
/* We figured there are NULL values in Brand & Product line in Transactions table, so we deal with those columns */
UPDATE transactions
SET brand='Others'
WHERE brand = '';

UPDATE transactions
SET product_line='Undefined'
WHERE product_line = '';

/* Cleaning the Gender column in the Customer Demographic table */
SELECT DISTINCT(gender) FROM cust_demographic;

UPDATE cust_demographic
SET gender='Female'
WHERE gender = 'F';

UPDATE cust_demographic
SET gender='Female'
WHERE gender = 'Femal';

UPDATE cust_demographic
SET gender='Male'
WHERE gender = 'M';

UPDATE cust_demographic
SET gender='Others'
WHERE gender = 'U';
```

1. The *gender* column in the ***cust_demographic*** has values which are wrongly spelled and inconsistent –

	gender character varying (512)
1	Femal
2	Male
3	M
4	U
5	Female

We obtain the below results:

	gender character varying (512)
1	Male
2	Others
3	Female

```
/* Cleaning the empty job titles */

SELECT COUNT(job_title) AS no_of_empty_job_titles FROM cust_demographic
WHERE job_title = '';

UPDATE cust_demographic
SET job_title='Not Updated'
WHERE job_title = '';

UPDATE new_cust_list
SET job_title='Not Updated'
WHERE job_title = '';

SELECT COUNT(job_title), job_title FROM cust_demographic
GROUP BY job_title
ORDER BY job_title;
```

2. Number of empty *job_title* in ***cust_demographic***

	no_of_empty_job_titles bigint
1	506

```
/* Cleaning the state column in Customer Address */
SELECT DISTINCT(state) FROM cust_address;

UPDATE cust_address
SET state='Queensland'
WHERE state = 'QLD';

UPDATE cust_address
SET state='Victoria'
WHERE state = 'VIC';

UPDATE new_cust_list
SET state='New South Wales'
WHERE state = 'NSW';
```

3. Cleaning the *state* column in ***cust_address & new_cust_list*** in order to maintain consistencies with the names.

	state character varying (512)
1	QLD
2	VIC
3	New South Wales
4	Victoria
5	NSW

Data results after cleaning:

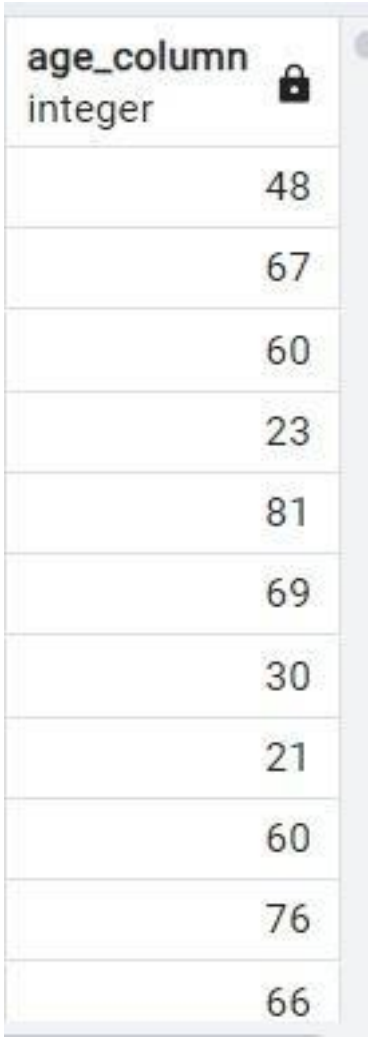
	state character varying (512)
1	New South Wales
2	Queensland
3	Victoria

```
/* Creating an Age column in Customer Demographics and New customer list*/
ALTER TABLE cust_demographic
ADD COLUMN age_column INTEGER;

UPDATE cust_demographic
SET age_column =
  CASE
    WHEN dob <> '' AND dob IS NOT NULL AND dob::date IS NOT NULL
    THEN EXTRACT(YEAR FROM AGE(dob::date))
    ELSE NULL
  END;

ALTER TABLE new_cust_list
ADD COLUMN age_column INTEGER;
```

```
UPDATE new_cust_list
SET age_column =
  CASE
    WHEN dob <> '' AND dob IS NOT NULL AND dob::date IS NOT NULL
    THEN EXTRACT(YEAR FROM AGE(dob::date))
    ELSE NULL
  END;
```



age_column integer
48
67
60
23
81
69
30
21
60
76
66

SQL PROBLEMS

```
/* Online orders by brand & Product line */
SELECT COUNT(online_order) AS number_of_online_orders, brand, product_line FROM transactions
WHERE online_order = 'TRUE'
GROUP by brand, online_order, product_line
ORDER by brand, product_line;
```

	number_of_online_orders bigint	brand character varying (512)	product_line character varying (512)
1	296	Giant Bicycles	Road
2	1263	Giant Bicycles	Standard
3	81	Giant Bicycles	Touring
4	115	Norco Bicycles	Mountain
5	442	Norco Bicycles	Road
6	860	Norco Bicycles	Standard
7	414	OHM Cycles	Road
8	1031	OHM Cycles	Standard
9	115	OHM Cycles	Touring
10	90	Others	Undefined
11	251	Solex	Road
12	1706	Solex	Standard
13	90	Solex	Touring
14	108	Trek Bicycles	Mountain
15	485	Trek Bicycles	Road
16	867	Trek Bicycles	Standard
17	76	WeareA2B	Road
18	1211	WeareA2B	Standard
19	328	WeareA2B	Touring

```sql
/* Top 10 customers with most number of transactions */
SELECT COUNT(ts.transaction_id) AS no_of_transactions, cd.customer_id FROM transactions ts
LEFT JOIN cust_demographic cd
ON ts.customer_id = cd.customer_id
GROUP BY ts.brand, cd.customer_id
ORDER BY COUNT(ts.transaction_id) desc
LIMIT 10;
```

	no_of_transactions bigint	customer_id integer
1	8	2783
2	7	171
3	7	1611
4	7	2353
5	6	1946
6	6	1451
7	6	2184
8	6	1591
9	6	528
10	6	2023

```sql
/* Segregation of users based on their Wealth Segment & comparison with total number of transactions*/
SELECT COUNT(DISTINCT(cd.customer_id)) AS count_of_customers,
   COUNT(ts.transaction_id) AS count_of_transactions,
   cd.wealth_segment
   FROM cust_demographic cd
LEFT JOIN transactions ts
ON ts.customer_id = cd.customer_id
GROUP BY wealth_segment
ORDER BY wealth_segment;
```

	count_of_customers bigint	count_of_transactions bigint	wealth_segment character varying (512)
1	979	4858	Affluent Customer
2	1021	5088	High Net Worth
3	2000	10051	Mass Customer

```sql
/* State wise distribution of Customers from Customer Address */
SELECT COUNT(customer_id) AS distribution_of_customers, state FROM cust_address
GROUP BY state
ORDER BY state;
```

	distribution_of_customers bigint	state character varying (512)
1	2140	New South Wales
2	838	Queensland
3	1021	Victoria

```sql
/* Age wise distribution of transactions of different cycle brands */
/*Based on the new age group created in the cust_demographic table (refer data_cleaning.sql), we carry out this analysis */
SELECT
  t.brand,
  CASE
    WHEN age_column< 18 THEN 'Under 18'
    WHEN age_column BETWEEN 18 AND 29 THEN '18-29'
    WHEN age_column BETWEEN 30 AND 39 THEN '30-39'
    WHEN age_column BETWEEN 40 AND 49 THEN '40-49'
    WHEN age_column BETWEEN 50 AND 59 THEN '50-59'
    WHEN age_column >= 60 THEN '60 and over'
    ELSE 'Unknown'  -- Handle any unexpected cases
  END AS age_group,
  COUNT(*) AS total_transactions
FROM transactions t
JOIN cust_demographic cd
ON cd.customer_id = t.customer_id
GROUP BY t.brand, age_group
ORDER BY t.brand, age_group;
```

	brand character varying (512)	age_group text	total_transactions bigint
1	Giant Bicycles	18-29	437
2	Giant Bicycles	30-39	546
3	Giant Bicycles	40-49	1027
4	Giant Bicycles	50-59	641
5	Giant Bicycles	60 and over	590
6	Giant Bicycles	Unknown	71
7	Norco Bicycles	18-29	381
8	Norco Bicycles	30-39	512
9	Norco Bicycles	40-49	949
10	Norco Bicycles	50-59	529
11	Norco Bicycles	60 and over	461
12	Norco Bicycles	Unknown	77
13	OHM Cycles	18-29	424
14	OHM Cycles	30-39	536
15	OHM Cycles	40-49	961
16	OHM Cycles	50-59	563
17	OHM Cycles	60 and over	495
18	OHM Cycles	Unknown	63
19	Others	18-29	28
20	Others	30-39	35
21	Others	40-49	62

	brand character varying (512)	age_group text	total_transactions bigint
22	Others	50-59	36
23	Others	60 and over	28
24	Others	Unknown	8
25	Solex	18-29	613
26	Solex	30-39	733
27	Solex	40-49	1332
28	Solex	50-59	808
29	Solex	60 and over	692
30	Solex	Unknown	74
31	Trek Bicycles	18-29	416
32	Trek Bicycles	30-39	519
33	Trek Bicycles	40-49	927
34	Trek Bicycles	50-59	547
35	Trek Bicycles	60 and over	512
36	Trek Bicycles	Unknown	69
37	WeareA2B	18-29	435
38	WeareA2B	30-39	561
39	WeareA2B	40-49	1041
40	WeareA2B	50-59	636
41	WeareA2B	60 and over	538
42	WeareA2B	Unknown	84

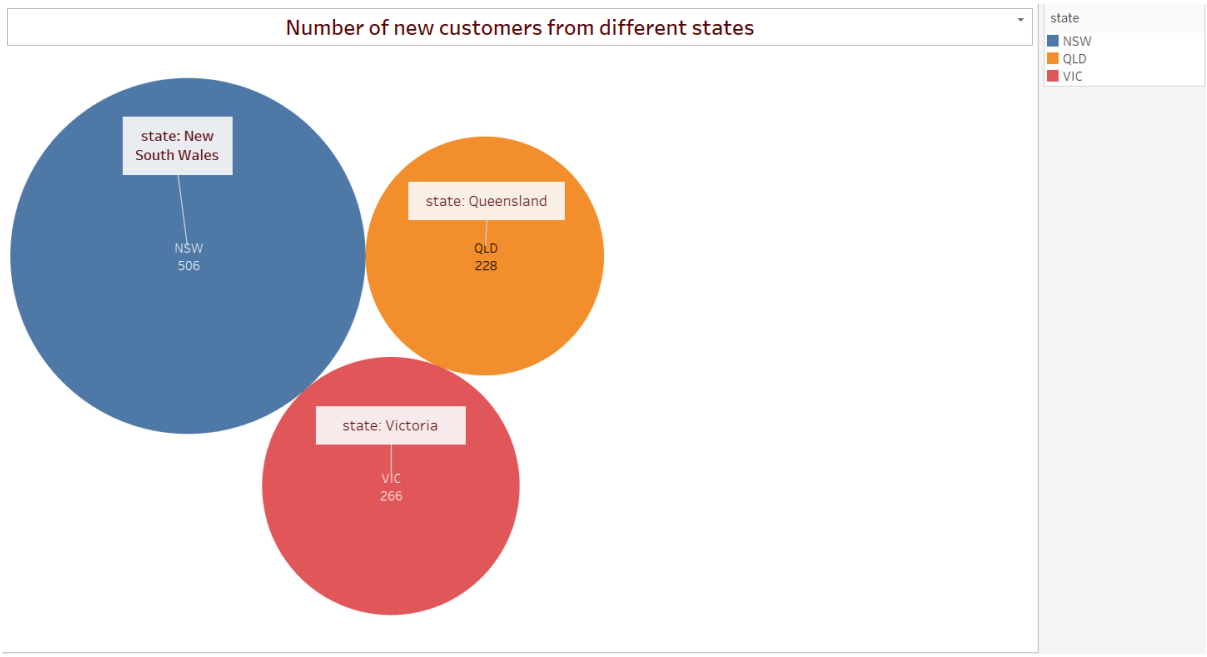
```sql
/* State wise and age group wise distribution of new users*/
SELECT state,
   CASE
      WHEN age_column< 18 THEN 'Under 18'
      WHEN age_column BETWEEN 18 AND 29 THEN '18-29'
      WHEN age_column BETWEEN 30 AND 39 THEN '30-39'
      WHEN age_column BETWEEN 40 AND 49 THEN '40-49'
      WHEN age_column BETWEEN 50 AND 59 THEN '50-59'
      WHEN age_column >= 60 THEN '60 and over'
      ELSE 'Unknown'  -- Handle any unexpected cases
   END AS age_group,
   COUNT(first_name) AS count_new_users
   FROM new_cust_list
GROUP BY state, age_group
ORDER BY state, age_group;
```

	state character varying (512)	age_group text	count_new_users bigint
1	New South Wales	18-29	70
2	New South Wales	30-39	47
3	New South Wales	40-49	104
4	New South Wales	50-59	90
5	New South Wales	60 and over	188
6	New South Wales	Unknown	7
7	Queensland	18-29	28
8	Queensland	30-39	26
9	Queensland	40-49	42
10	Queensland	50-59	42
11	Queensland	60 and over	88
12	Queensland	Unknown	2
13	Victoria	18-29	36
14	Victoria	30-39	34
15	Victoria	40-49	61
16	Victoria	50-59	51
17	Victoria	60 and over	76
18	Victoria	Unknown	8

```
/* Maximum and Minimum Transaction Value */
SELECT CONCAT('$',MAX(list_price)) AS max_list_price, CONCAT('$',MIN(list_price)) AS min_list_price FROM
transactions;
```

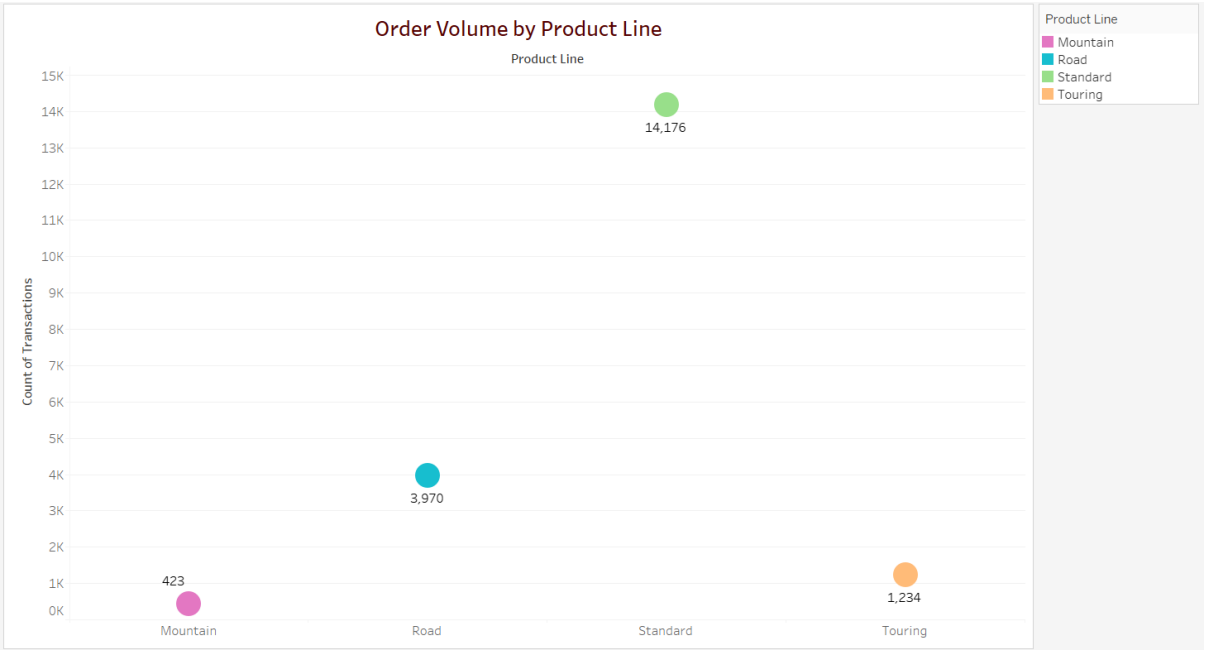
	max_list_price text	min_list_price text
1	$2091.47	$12.01

VISUALIZATIONS

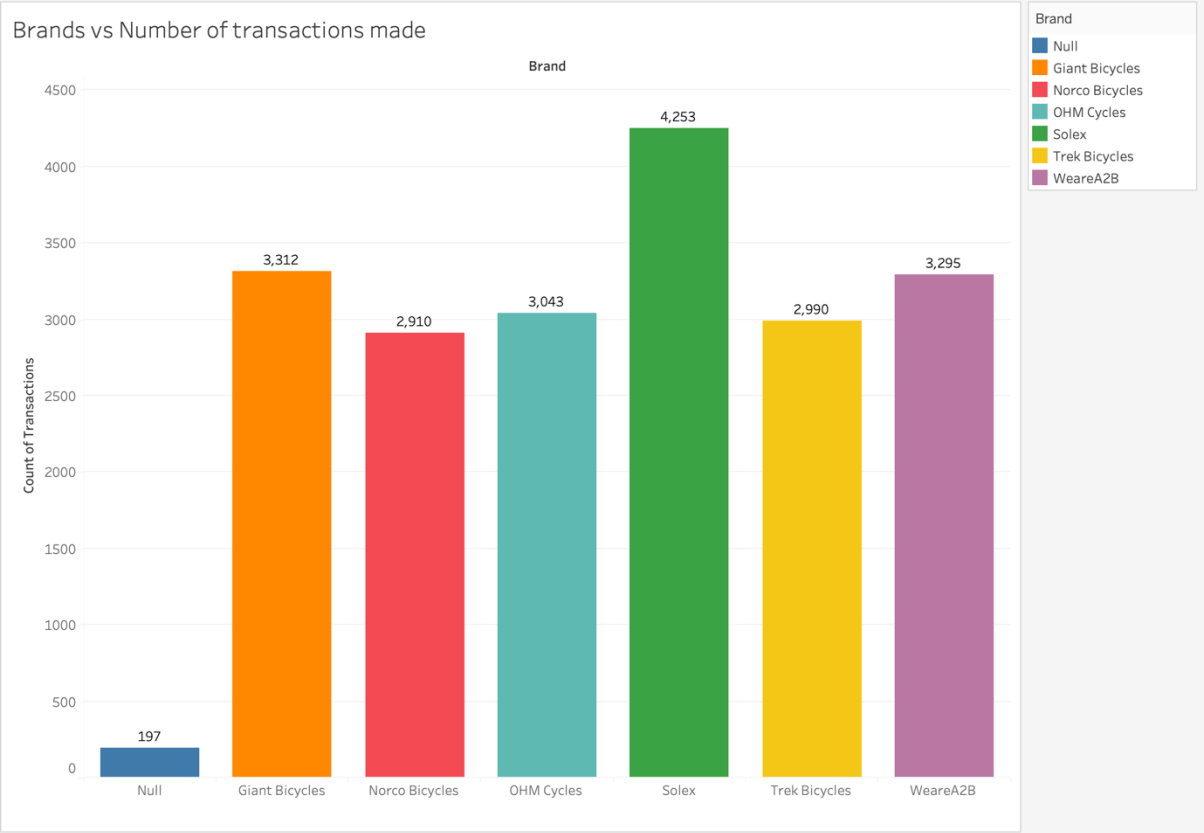


This visualization gives a summary of the number of new customers from different states we are taking into consideration. We can investigate how the customers are divided across the states and see that New South Wales has the maximum number of new customers and both Queensland and Victoria have equal number of new customers. Based on this analyzation we can see how the production of bikes should be in the respective states.

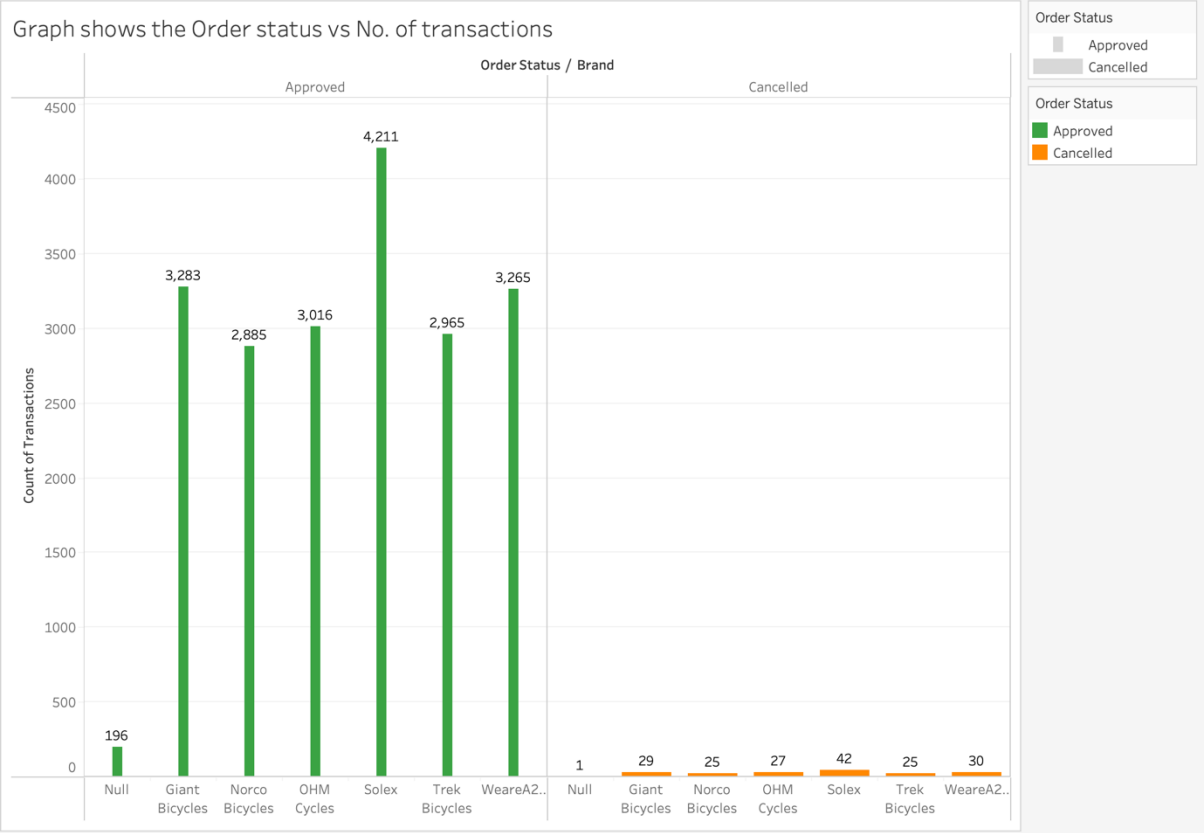
This visualization gives a summary of the number of new customers from different job industries we are taking into consideration. We can investigate how the customers are divided across the industries and see that Financial Services has the maximum number of new customers and both agriculture and telecommunications have the least number of new customers. Based on this we can analyze how the marketing has to be done in which section in order to increase the number of bike sales overall.



This visualization gives an analysis from a overall transaction perspective. It shows the total order volume by the product line. It shows how many transactions have been done for each product line. One would be able to analyse where the maximum profit has been and see where the minimum profit is also. By knowing this the production department would know where they have been lacking and what can be done to improve and gain more profits.



The number of transactions broken down per brand is displayed in the above graph. As we can see, the brand with the most transactions was Solex. Following the Solex brand, Giant bicycles have generated the most transactions, while Norco bicycles have generated the fewest.



The graphic illustrates how we are determining the number of approved and cancelled transactions based on the brands. It is evident that the Solex brand has authorized a maximum number of transactions, and it has also cancelled some transactions. Norco Bicycles has allowed the fewest transactions, while Trek Bicycles has cancelled the fewest transactions.