# Bee Cycle

A SQL Analysis of a Cycling Company

Lara Carter

# Table of Contents

# Project Summary

# Project Summary

**About:** Bee Cycle is a fictitious cycling company that opened in July 2017 and sells a range of cycling products in stores across the globe.

**Problem:** Since the COVID-19 pandemic, Bee Cycle has started to suffer a decline in profits.

**Objective:** Bee Cycle has asked you, a data analyst, to provide data driven insights to help advise the marketing team on how to increase profits for the company.

| Skills Gained | Tools Used |
| --- | --- |
| ➢ Sorting and Filtering<br>➢ Data Cleaning and Wrangling<br>➢ Exploratory Data Analysis<br>➢ Queries, subqueries, and common table expressions<br>➢ Joining data from relational datasets | BigQuery  |
| ➢ Data Visualisation | Tableau  |

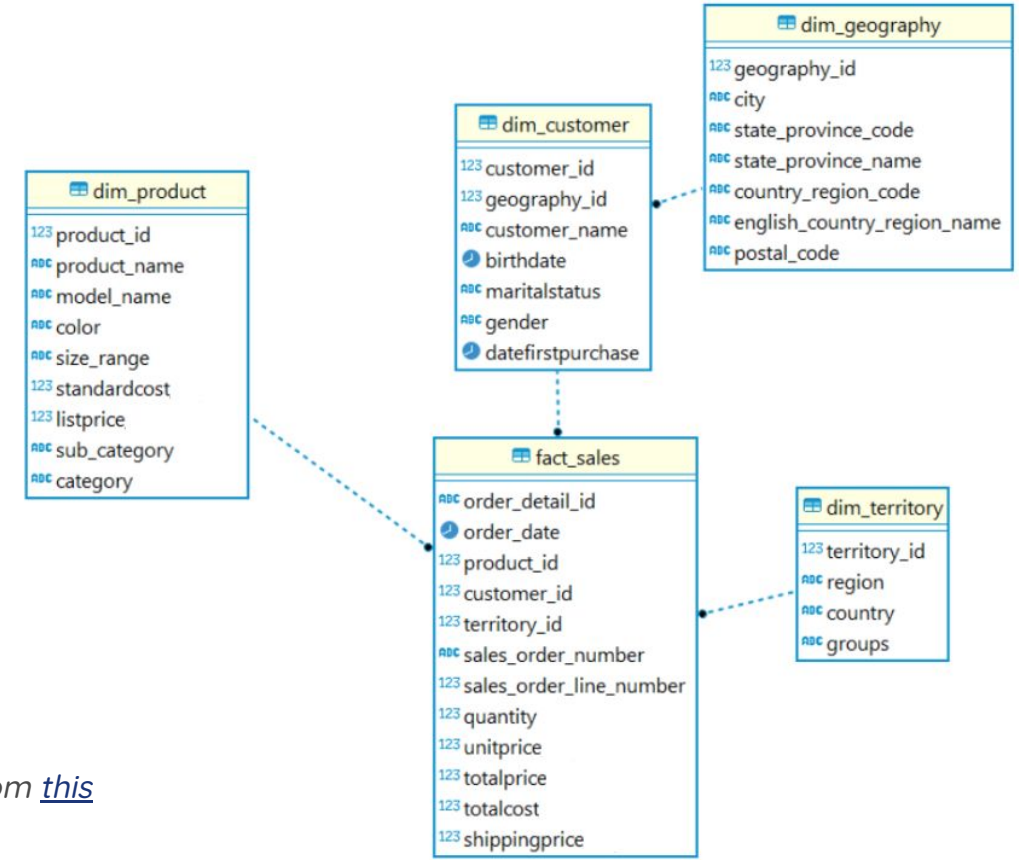
# Data

# Data

Bee Cycle provided **5 datasets*** to analyse.

Their relations are summarised in this entity relationship diagram:

A data dictionary is available to view here.




*All data is fictitious and was retrieved from this public repository.

# Evaluation of the Data

- The dataset is **comprehensive** in that it includes information on a range of variables including locations, demographics, product colours and sizes, and transactions since store opening.
  - However, additional granularity such as the time an order was placed, whether purchases were in-store or online, or more detailed product features may provide further insights to help answer the business questions.

- The data can be considered **reliable**, in that it is first-party data that is complete and consistent and generally free from outliers, duplicates or biases.
  - However, the data is **not wholly up to date**, as it only provides sales data until June 2022. More current data could therefore further enhance reliability.

# Project Process

# Project Process

1. **Upload data (5 CSV files) to BigQuery**

2. **Initial inspection of the data**
   - ➢ Assessed each variable, checked their data types, and number of rows/columns
   - ➢ Assessed table relations and primary and foreign keys
   - ➢ Created a data dictionary

# Project Process

3 **Data cleaning, wrangling and analysis using SQL queries**

- ➢ Sorted and filtered the data
- ➢ Checked and handled missing, null, outlying and erroneous values, as well as any spelling errors or misfielded values.
- ➢ Used queries such as aggregate, window, ranking, string, conversion, conditional, and date and time functions, as well as join statements.
- ➢ Used subqueries and common table expressions to answer complex questions.

4 **Create data visualisations in Tableau**

# Analysis

# Key Questions

We will focus our analysis on the following questions:

1) **Which products are most profitable?**
2) **Which months are most profitable?**
3) **Which regions are most profitable?**
4) **What type of customer generates the most profit?**
5) **Which customers spend the most money?**
6) **a. Which cities are most profitable?**
   **b. What are the most popular products within these cities?**
   **c. In what colours?**

# Data Overview

**1360** customers

**119** products

**6** countries

**5954** orders

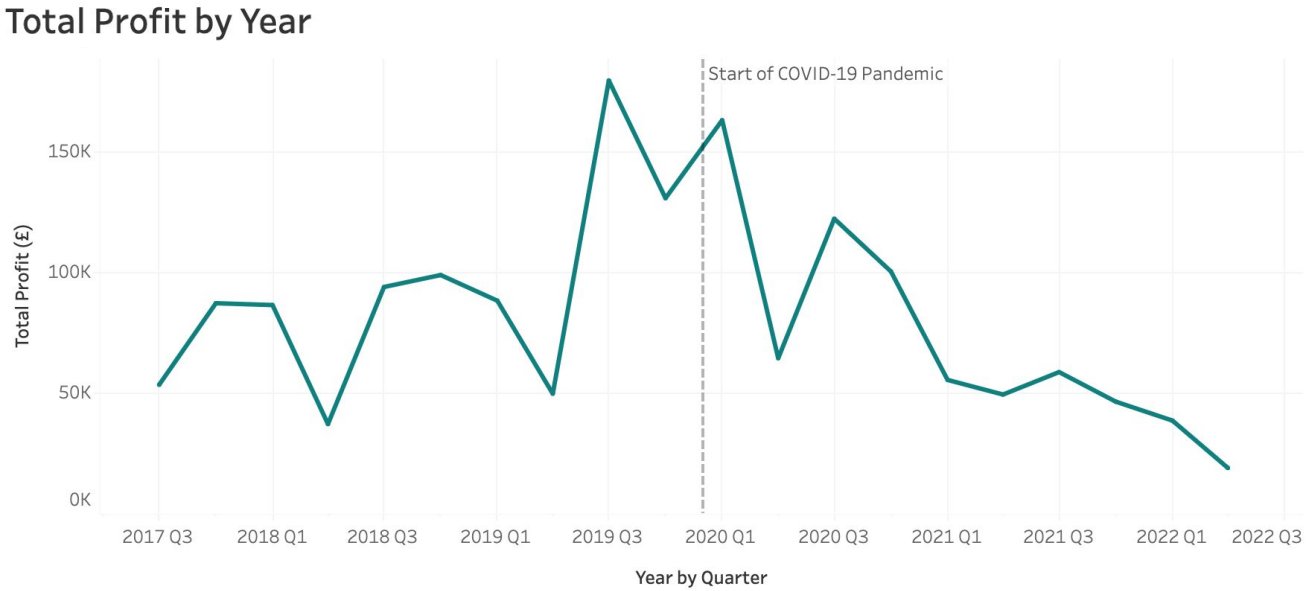
**£3,862,950.25**

total revenue

**£1,621,608.21**

total profits

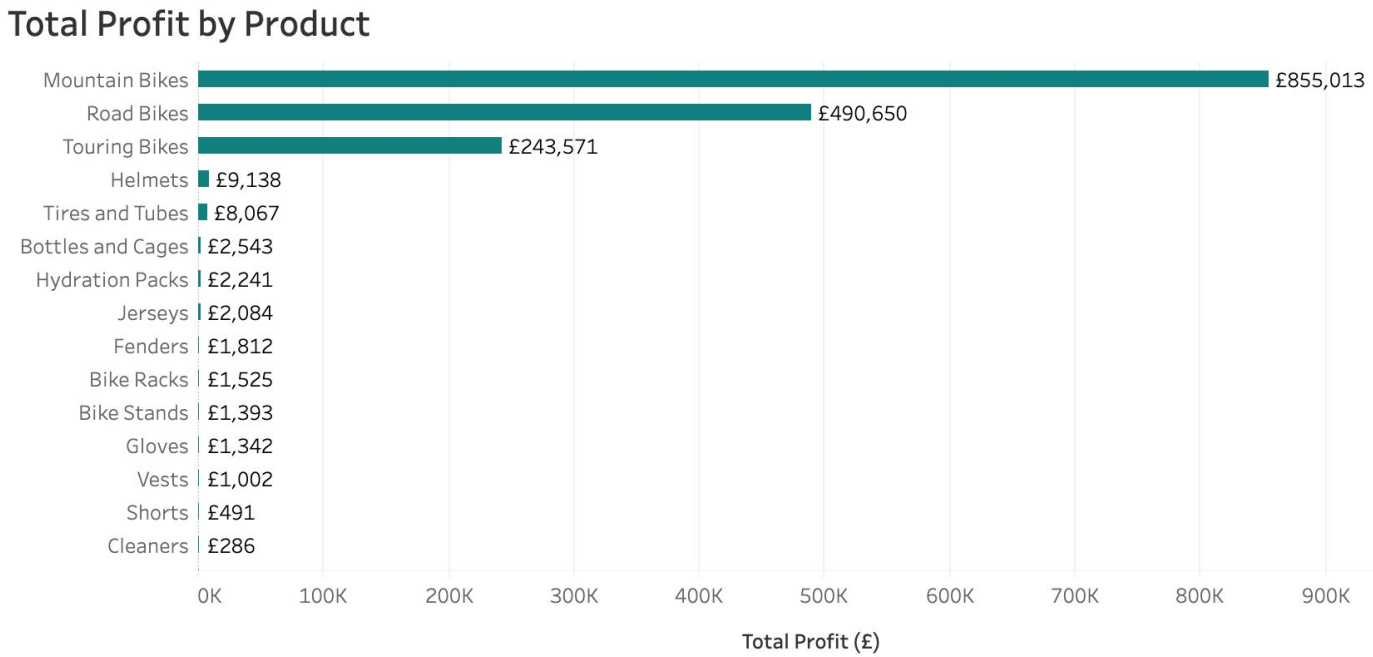
# Data Overview



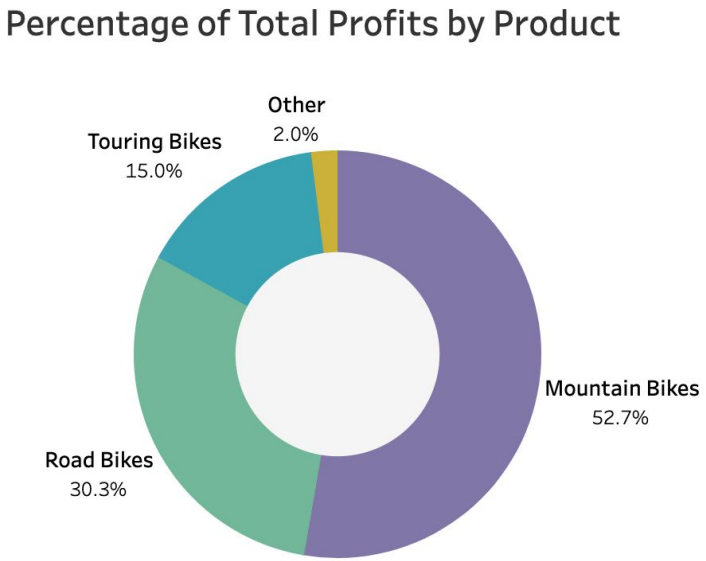

There was a sharp decline in profits shortly after the COVID-19 pandemic began in Dec 2019.

Profits briefly increased but have **continued to decline again since the end of Q3 2020**.

# Most Profitable Products



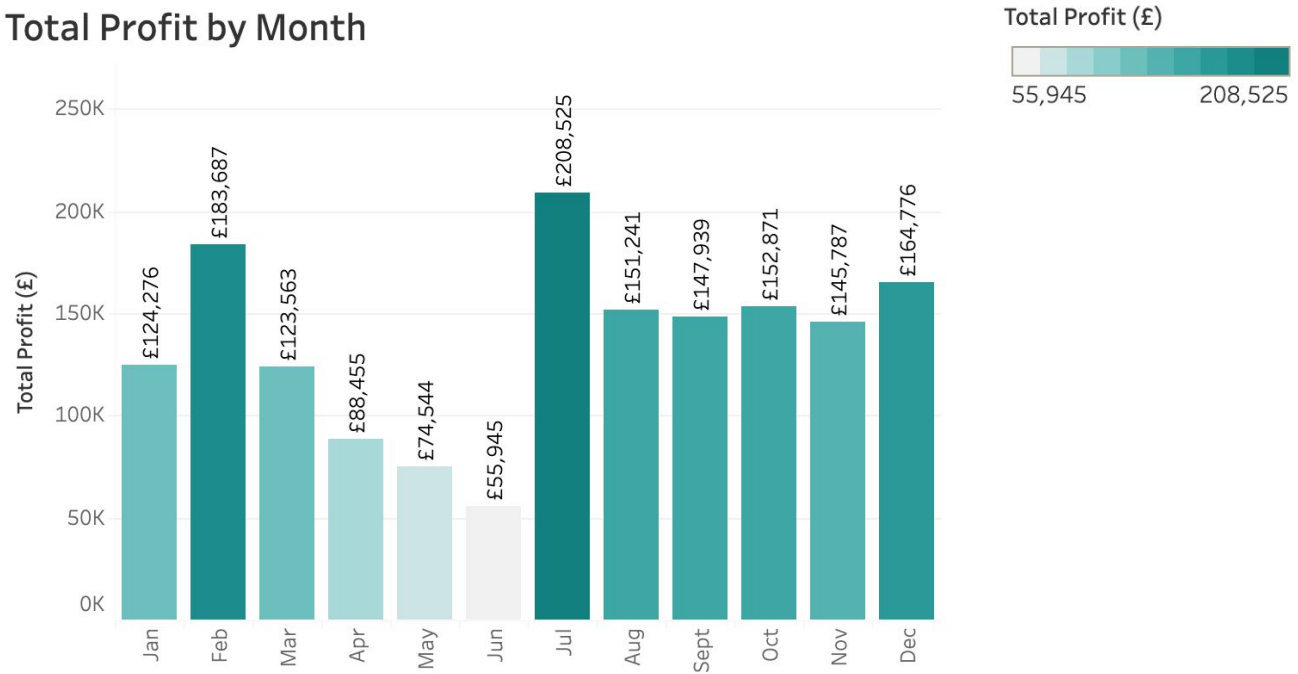
# Most Profitable Products



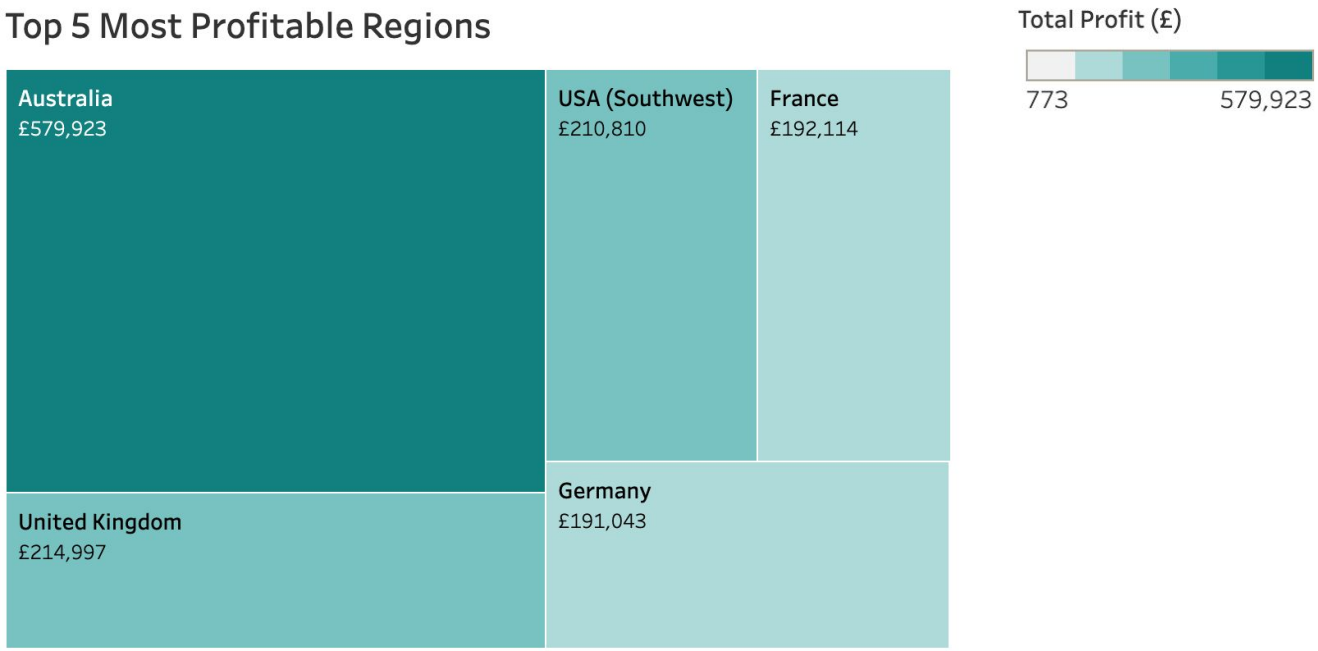

As seen from the current and previous slide, the most profitable products are:

1) **Mountain Bikes**
2) **Road Bikes**
3) **Touring Bikes**

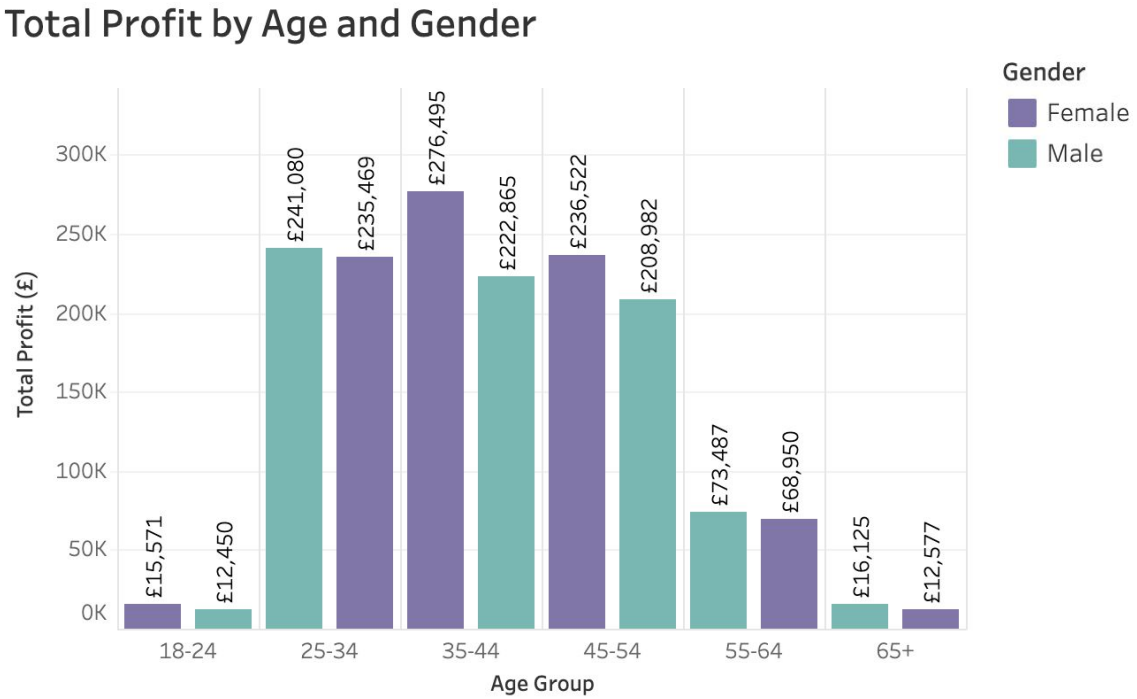
# Most Profitable Months



# Top 5 Most Profitable Regions



# Most Profitable Customer Type



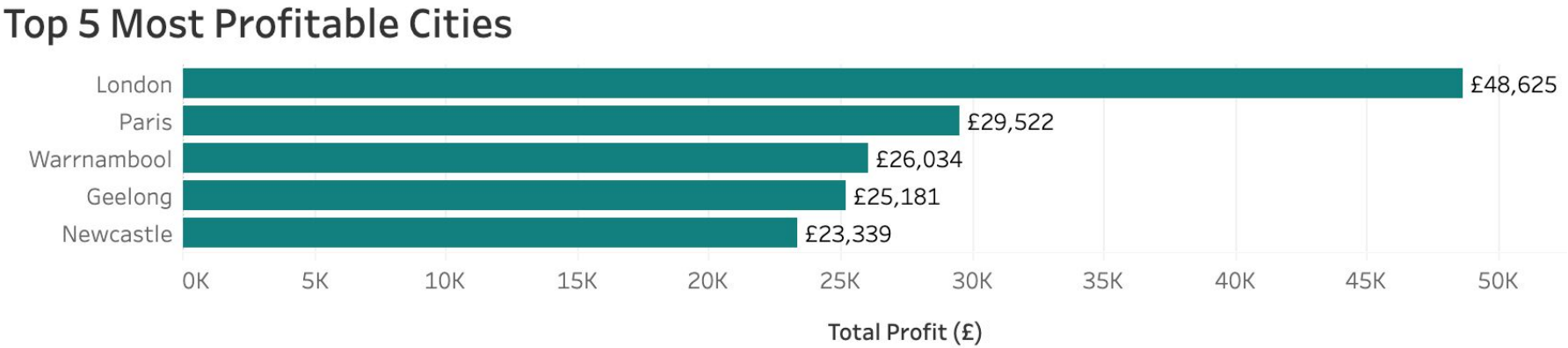

Customers generating the most profit are **men/women aged 25-54.**

The single most profitable customer type is **women aged 35-44.**

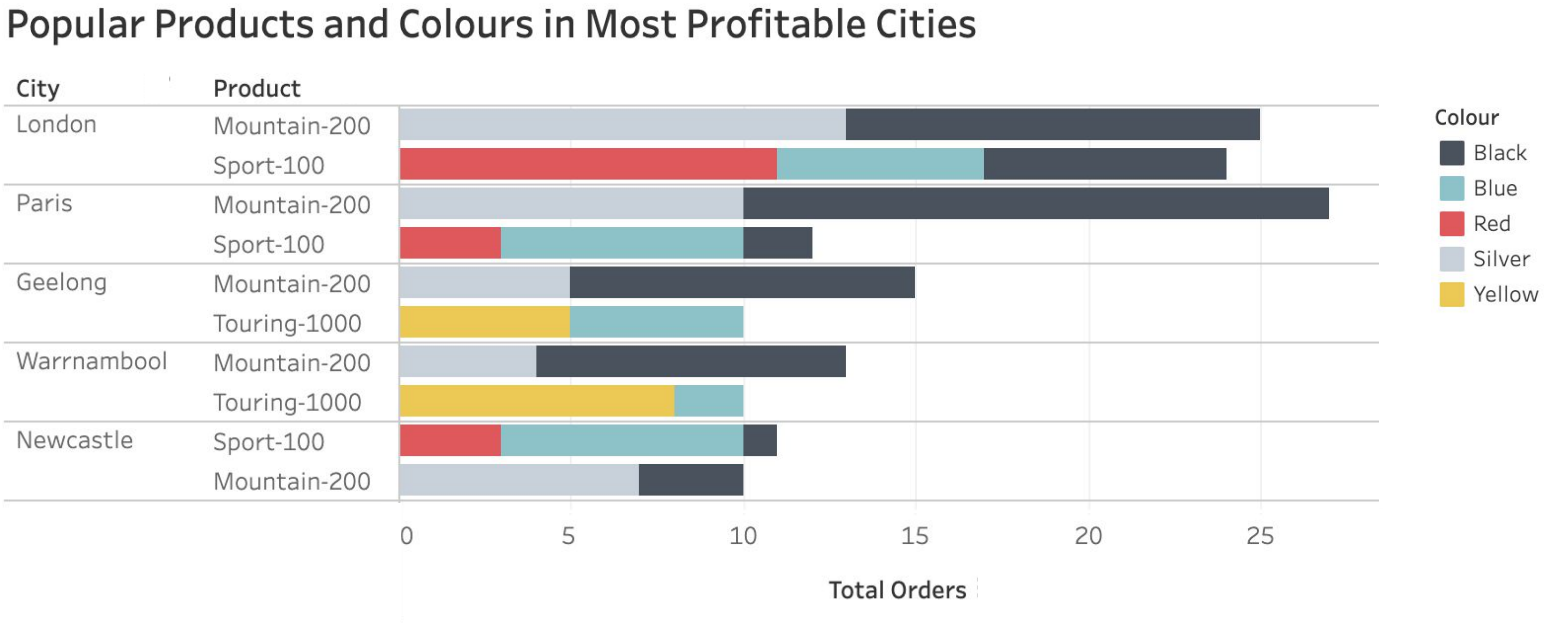
# Top 10 Highest Spending Customers

| Name | Total Spent |
|---|---|
| Nichole Nara | £9307 |
| Kaitlyn Henderson | £9306 |
| Margaret He | £9288 |
| Randall Dominguez | £9286 |
| Adriana Gonzalez | £9270 |
| Rosa Hu | £9251 |
| Brandi Gill | £9237 |
| Brad She | £9221 |
| Francisco Sara | £9215 |
| Kate Anand | £7610 |

# Top 5 Most Profitable Cities



# Top 2 Most Popular Products & Colours in Profitable Cities



# Actions

# Key Recommendations

1. Focus marketing on promoting products that generate the most profit: **mountain bikes, road bikes, and touring bikes**.

2. Schedule marketing campaigns during **July** (the most profitable month). Consider sponsoring cycling events and competitions that take place during July such as the **Tour de France.**

3. Increase marketing in regions that generate the most profit: **Australia, UK, Southwest USA, France and Germany.**

# Key Recommendations

4 Target marketing towards customer profiles that generate the most profit: **men/women aged 25-54** and specifically **women aged 35-44**.

5 Consider implementing a **loyalty scheme** that provides rewards and discounts to high spenders. This will maintain good relations and encourage spending.

6 Develop billboards in the top 5 most profitable cities: **London, Paris, Warrnambool, Geelong, and Newcastle**.

Ensure billboards display the top 2 most popular products for that particular city and in the most popular colours.

# Next Steps

- Monitor and evaluate the effectiveness of implemented recommendations. For example, assess whether profits have increased since loyalty schemes were introduced.

- Gather additional data sources, such as more up-to-date sales data, customer feedback, product features, and competitor information, to enrich the analysis and gain deeper insights into customer behavior and market dynamics.

# View SQL Queries on GitHub here