BIKE FEATURE ANALYSIS USING SQL

Introduction

The Bike Features Dataset is a comprehensive collection of data on various bike models, encompassing essential specifications and features that define each bike's performance, design, and usability. This dataset has been compiled to provide insights into the diverse characteristics of bikes from different brands, making it a valuable resource for consumers, automotive analysts, and enthusiasts.

The Bike Features Dataset provides a rich source of information that allows for in-depth analysis and

comparison of different bike models. By analysing this data, one can explore relationships between features, such as how engine displacement and torque relate to fuel efficiency and on-road price. Additionally, this dataset enables brand-level analysis to uncover each manufacturer's strengths and areas of expertise, assisting consumers in making informed purchase decisions. Whether the goal is to find the best bike for city commuting, long-distance touring, or performance racing, this dataset offers the critical data points needed for thorough evaluation and selection.

Objectives

The primary objective of this analysis is to extract valuable insights from the Bike Features Dataset that will enable potential buyers, automotive analysts, and manufacturers to better understand the distinguishing characteristics of various bike models and variants. By analysing critical specifications such as engine performance, fuel efficiency, pricing, and design attributes, this project aims to support informed decision-making for consumers looking to purchase a bike that aligns with their needs and preferences. Additionally, the insights derived from this dataset can aid

manufacturers in identifying market trends, understanding consumer preferences, and enhancing product development. For analysts, the data offers a rich source for comparative studies and market research, further contributing to an informed and competitive automotive industry.

Overview of Dataset

- **Rows:** 543
- **Columns:** 14

Column Name	Description
Variant Name	Version of the bike model with specific features.
Model_name	Main model name of the bike.
on_road_price	Purchase price, including taxes and fees.
Engine Type	Engine configuration (e.g., single-cylinder, 4-stroke).
Displacement	Engine capacity in cc, affecting power.
Max Torque	Peak torque output in Nm.
Cooling System	Engine cooling type (e.g., air-cooled, liquid-cooled).
start_type	Engine start method (e.g., kick, self-start).
Fuel Supply	Fuel delivery system, such as fuel injection.
Ignition	Ignition type for combustion control.

Gear Box	Number of gears (e.g., CVT, 6-speed).
Mileage	Fuel efficiency in km/l.
Body Type	Classification by bike design (e.g., scooter, cruiser).
company	Manufacturer or brand name.

Data Cleaning and Preparation

To optimize the dataset for accurate analysis, various data cleaning and preparation steps were implemented to remove inconsistencies, enhance clarity, and organize the data for optimized querying efficiency.

- **Removing Unnecessary Columns:** Dropped columns irrelevant to this analysis.
- **Data Type Standardization:** Values in columns like stroke, city mileage, and displacement were standardized by removing units (e.g., "kmpl" and "cc") for consistency.
- **Column Renaming:** Some columns were renamed for clarity (e.g., city mileage to Mileage and company name to Model_name).

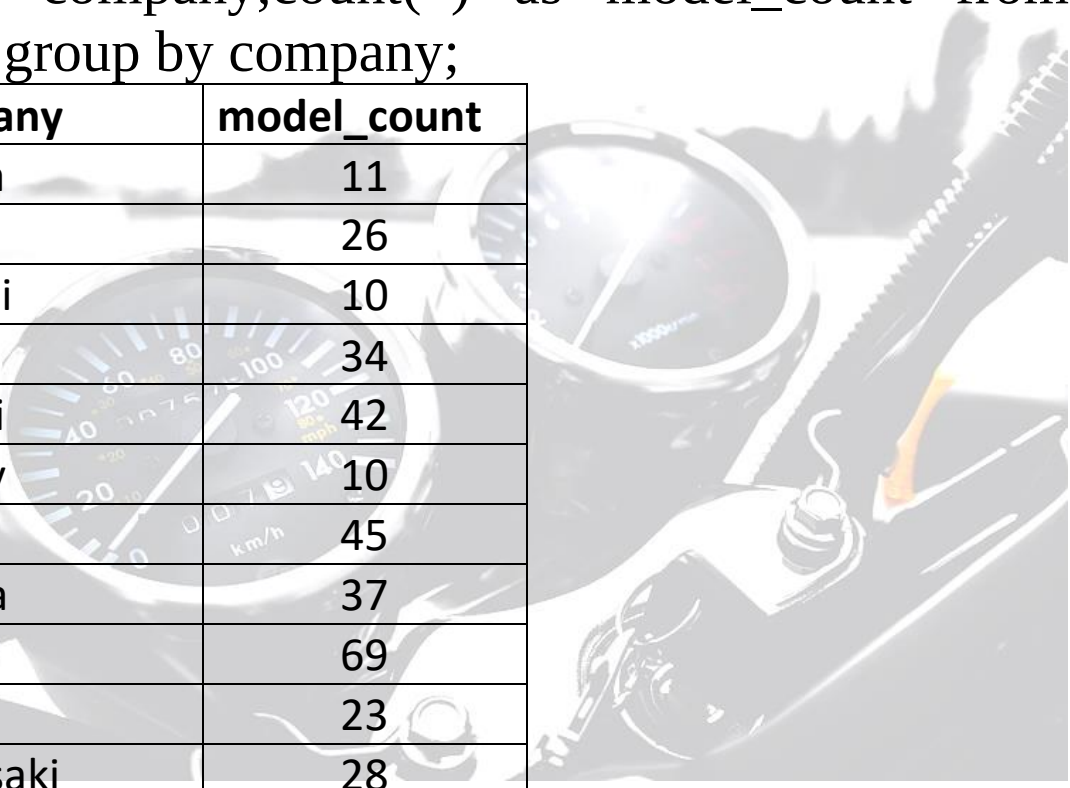
- **New Column Addition:** A company column was added to store the brand extracted from variant name.

Analysis:

Q. Count the number of bikes models per brand:

select company,count(*) as model_count from bikes group by company;

company	model_count
Aprilia	11
Bajaj	26
Benelli	10
BMW	34
Ducati	42
Harley	10
Hero	45
Honda	37
Indian	69
Jawa	23
Kawasaki	28
KTM	11
Royal	43
Suzuki	26
Triumph	35
TVS	48
Yamaha	26
Yezdi	13



Q. Get average mileage for each vehicle type:

```
select `variant name`,avg(mileage) as avg_mileage from bikes group by `variant name`;
```

variant name	avg_mileage
Aprilia SR 160 Carbon	35
Aprilia SR 160 STD	35
Aprilia SR 160 Race	35
Aprilia RS 660 STD	20.4
Aprilia SR Storm STD	35
Aprilia RSV4 Factory	15.4
Aprilia SXR 125 STD	40
Aprilia SXR 160 STD	35
Aprilia Tuono V4 Factory	14
Aprilia Tuono 660 STD	20.4

Q. List bikes that use Fuel Injection as the fuel system:

```
select `variant name`,`fuel supply` from bikes where `fuel supply` like '%Fuel Injection';
```

variant name	fuel supply
Aprilia RS 660 STD	Fuel Injection
Aprilia RSV4 Factory	Fuel Injection
Aprilia SXR 125 STD	Fuel Injection
Aprilia SXR 160 STD	Fuel Injection
Aprilia Tuono V4 Factory	Fuel Injection
Aprilia Tuono 660 STD	Fuel Injection
Aprilia SR 125	Fuel Injection

Bajaj Pulsar N160 Dual Channel ABS	Fuel Injection
Bajaj Pulsar 220 F STD	Fuel Injection
Bajaj Pulsar NS200 2023 STD	Fuel Injection

Q. Get the model names and prices of motorcycles with a mileage greater than 30:

select Model_name,`On-road prize`,mileage from bikes where mileage >30;

Model_name	On-road prize	mileage
Piaggio	147307	35
Piaggio	144523	35
Piaggio	154779	35
Piaggio	124892	35
Piaggio	147494	40
Piaggio	158535	35
Piaggio	134660	40
Bajaj	156165	59.11
Bajaj	164029	40
Bajaj	176851	40.36

Q. Find motorcycles that are Sports Bikes :

select `variant Name`,`body Type` from bikes where `body type` like 'SP%';

variant Name	**body Type**
Aprilia RS 660 STD	Sports Bikes
Aprilia SR Storm STD	Sports Bikes

Aprilia RSV4 Factory	Sports Bikes
Aprilia Tuono V4 Factory	Sports Bikes
Aprilia Tuono 660 STD	Sports Bikes
Bajaj Pulsar N160 Dual Channel ABS	Sports Bikes
Bajaj Pulsar 220 F STD	Sports Bikes
Bajaj Pulsar NS200 2023 STD	Sports Naked Bikes, Sports Bikes
Bajaj Dominar 400 STD	Sports Naked Bikes, Sports Bikes
Bajaj Pulsar NS160 2023 STD	Sports Bikes

Q. Select all bikes with an engine capacity above 200cc:

select `Variant name`,`engine type`,displacement from bikes where displacement>200;

Variant name	**displacement**
Aprilia RS 660 STD	660
Aprilia RSV4 Factory	1099
Aprilia Tuono V4 Factory	1100
Aprilia Tuono 660 STD	660
Bajaj Pulsar 220 F STD	220
Bajaj Dominar 400 STD	373.3
Bajaj Pulsar F250 All-Black	249.07
Bajaj Avenger 220 Street	220
Bajaj Avenger Cruise 220 BS6	220
Bajaj Pulsar N250 All-Blackz	249.07

Q.select company,avg(mileage) as Avg_mileage from bikes group by company order by avg_mileage desc limit 5;

company	Avg_mileage
Hero	83.99777778
Honda	63.01756757
TVS	57.40229167
Yamaha	55.98692308
Bajaj	49.85730769

Q. calculates the range and count of On-road prize by model name.

select `Model_name`, min(`On-road prize`) as Min_Price, max(`On-road prize`) as Max_Price, count(*) as Model_Count from bikes group by `Model_name` order by Model_Count desc;

Model_name	Min_Price	Max_Price	Model_Count
Indian Motorcycyle	1939512	5296744	69
TVS Motors	53200	307024	48
Ducati Motor Holding	1046694	7756590	45
Hero Motocorp	84850	175572	45
Eicher Motors	173454	439398	43
Honda	77969	4347795	37

Triumph Motorcycles Ltd	273134	2416088	35
BMW Motorrad	321282	6099155	34
Kawasaki Motorcycle & Engine Company	200107	8852333	28
Bajaj	81970	274687	26
Suzuki Motorcycle India	95884	1886460	26
Yamaha Motor India	96030	228446	26
JAWA Motorcycles	192391	255474	23
Harley-Davidson	268751	4494793	13
Yezdi Motorcycle	234789	250200	13
Piaggio	124892	2731987	11
KTM Asia Motorcycle Manufacturing, Inc.	204714	419536	11
Qianjiang Motorcycle	275326	713603	10

Bivariate Analysis:

Q. Compare the average price of bikes based on the number of gears in the Gear Box

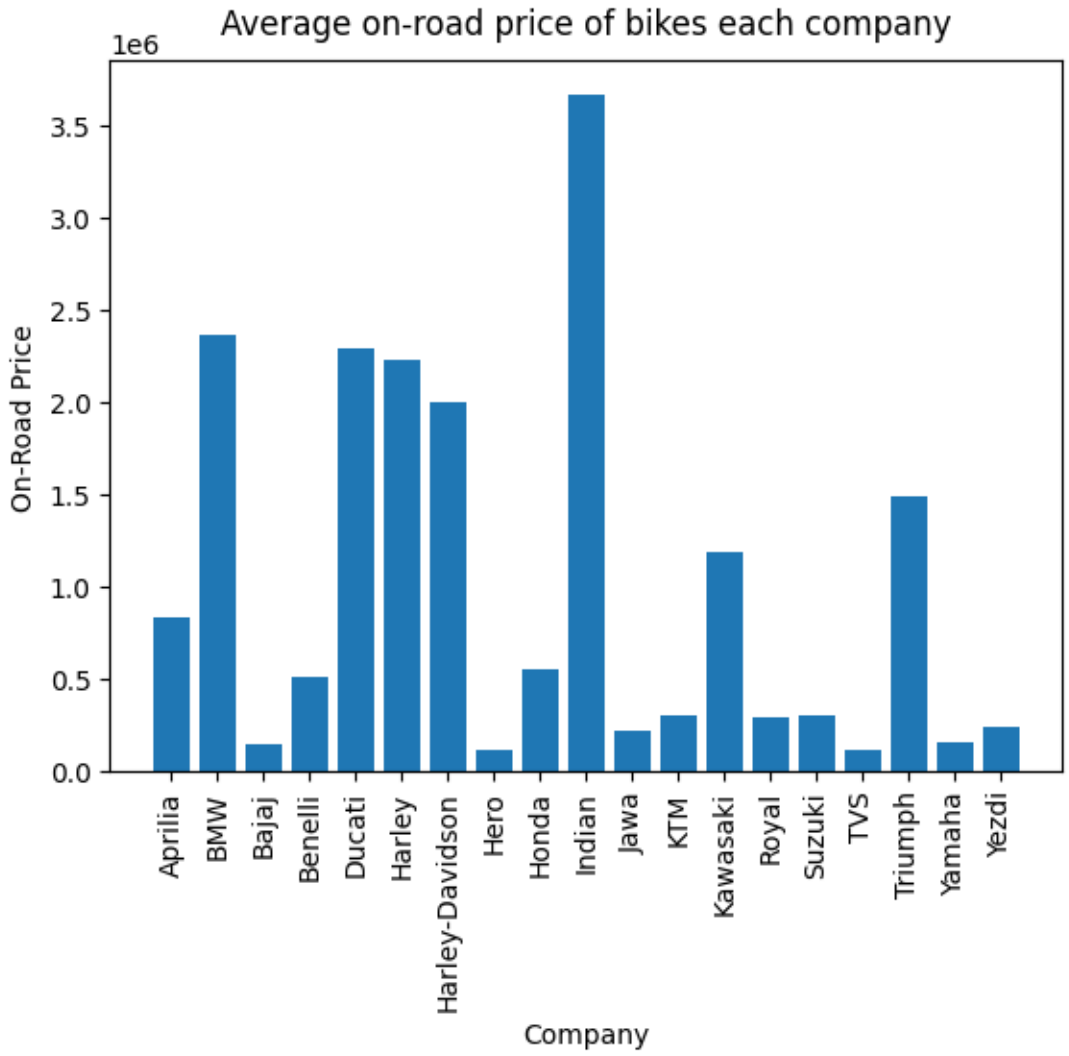
Gear Box	Avg_Price	Count
7 Speed	4347795	1
6 speed with Ducati Quick Shift up/down	2466876	1
Automatic Dual Clutch Transmission	1948978	1
6 Speed	1858202.854	309

Manual Transmission	1779086	1
6-speed, return	1208580	2
5-speed	1085043.5	2
5-Speed Return	943955.5	2
6 Speed Constant Mesh	817682	6
5 Speed	233477.8636	110
5 Speed constant mesh	150871.5862	29
V-Belt Automatic	116664.8182	11
CVT	108557.5897	39
Variomatic Drive	95281.25	4

Q. Analyze the average and range of On-road prize for each type of Cooling System

Cooling System	**Avg_Price**	**Min_Price**	**Max_Price**
Water/oil-cooled	3705327.2	2118656	6099155
Air-Liquid Cooled	2203426	1814590	2505649
Liquid Cooled	1688390.6	172856	8852333
Air Cooled	803939.72	53200	5230495
Air & Oil Cooled	771219.9737	100139	3500774
Oil Cooled	668277.1944	115550	3666628
NA- Electric Vehicle	153601	153601	153601

Q.Visualize the average on-road price for each bike company



Conclusion:

The dataset contains detailed information on various bike models, covering attributes such as on-road price, engine type, displacement, max torque, cooling system, start type, fuel supply, ignition type, gearbox, mileage, body type, and manufacturer. Here are the key findings:

1. **Price Range and Model Diversity**:
    - The dataset spans a wide range of on-road prices, from approximately 53,200 to 8,852,333 INR. This range reflects a diverse selection of bike models, from affordable scooters to high-end sports and cruiser bikes
    - With over 500 unique models, this dataset provides a comprehensive view of various segments, which could assist in market segmentation analysis and pricing strategies.
2. **Engine Specifications**:
    - The engine displacement varies significantly, from as low as 64cc to as high as 2458cc, indicating a variety of performance levels suited to different user needs—from city commuting to high-speed travel.
    - Most bikes are equipped with single-cylinder engines, though larger models include multi-cylinder configurations, especially in premium or sports categories.
3. **Fuel Efficiency and Mileage**:
    - The average mileage is approximately 39.7 km/l, with a wide standard deviation,

indicating significant variation in fuel efficiency across different models. This could help customers choose bikes based on fuel economy preferences.

4. **Manufacturer Presence**:
    - The dataset includes entries from 20 different companies, with brands like Indian Motorcycles having a significant presence. This can be useful for analysing brand market share and popularity in various bike segments.
5. **Target Market Insights**:
    - With a mix of scooters, sports bikes, and cruisers, this dataset covers bikes suitable for urban commuting, adventure touring, and leisure riding. Such data can aid manufacturers and dealers in understanding market trends and the preferences of different consumer segments.