Case Study

Rental Bike Sharing Data Analysis

Muhammad Fadli Syukur

Introduction

Studi kasus ini menganalisis dataset rental bike sharing yang didapatkan dari Kaggle, dengan memanfaatkan Google Sheet untuk proses persiapan data, menggunakan SQL Server untuk proses analisis, dan menggunakan Tableau Public untuk visualisasi data. Agar data dapat disajikan secara informatif yang memungkinkan untuk melihat insight dalam data dan memudahkan proses pengambilan keputusan.

About Dataset

Context

Sistem berbagi sepeda adalah generasi baru dari persewaan sepeda tradisional di mana seluruh proses mulai dari keanggotaan, penyewaan, dan pengembalian menjadi otomatis. Melalui sistem tersebut, pengguna dapat dengan mudah menyewa sepeda dari posisi tertentu dan mengembalikannya pada posisi lain. Saat ini, terdapat lebih dari 500 program berbagi sepeda di seluruh dunia yang terdiri dari lebih dari 500 ribu sepeda. Saat ini, terdapat minat yang besar terhadap sistem ini karena peran pentingnya dalam masalah lalu lintas, lingkungan dan kesehatan.

Terlepas dari penerapan sistem bike sharing di dunia nyata yang menarik, karakteristik data yang dihasilkan oleh sistem ini menjadikannya menarik untuk penelitian. Berbeda dengan layanan transportasi lain seperti bus atau kereta bawah tanah, durasi perjalanan, posisi keberangkatan dan kedatangan dicatat secara eksplisit dalam sistem ini. Fitur ini mengubah sistem bike sharing menjadi jaringan sensor virtual yang dapat digunakan untuk mendeteksi mobilitas dalam kota. Oleh karena itu, diharapkan sebagian besar peristiwa penting di kota dapat dideteksi melalui pemantauan data ini.

Content

I used dataset consisting of 17379 rows and 17 columns, here is a description of each column:

- instants : record index
- dteday : date
- season : season (1: winter, 2: spring, 3: summer, 4: fall)
- yr : year (0: 2011, 1: 20212)
- mnth : month (1 to 12)
- Hr : hour (0 to 23)
- Holiday : weather day is holiday or not
- Weekday : day of the week
- Workingday : if day is neither weekend nor holiday is 1, otherwise is 0

- weathersit:
  - 1 : clear, few clouds, partly cloudy
  - 2 : mist + cloudy, mist + broken clouds, mist + few clouds, mist
  - 3 : light snow, light rain + thunderstorm + scattered clouds, light rain + scattered clouds
  - 4 : heavy rain, ice pallets, thunderstorm, mist, snow + fog
- Temp : normalized feeling temperature in celsius. The values are derived via (t-t_min)/(t-t_max), t_min = -8, t_max = +39 (only in hourly scale)
- Atemp : normalized feeling temperature in celsius. The values are derived via (t-t_min)/(t-t_max), t_min = -16, t_max = +50 (only in hourly scale)
- Hum : normalized humidity, the values are divided to 100 (max)
- windspeed : normalized wind speed, the values are divided to 67 (max)
- Casual : count of casual users
- Registered : count of registered users
- Cnt : count of total rental bikes including both casual and registered

Data Preparation

Sebelum data di analysis saya melakukan pembersihan data. Dengan melakukan proses ini kita dapat menjaga kualitas data, mudah dalam melakukan interpretasi, memudahkan visualisasi dan sebagainya. Proses pembersihan data dilakukan menggunakan Google Sheet, berikut adalah tahapannya:

1. Mengecek nilai duplikasi dan missing value.
2. Menghapus kolom ‘mnth’ dan ‘day’, karena untuk nilai ini saya bisa memanfaatkan kolom ‘dteday’.
3. Mengganti nilai dari beberapa kolom kategorikal untuk mempermudah proses analysis dan visualisasi. (misalnya: kolom ‘season’ dibuah nilainya menjadi winter, spring, summer dan fall).
4. Mengganti nama kolom agar lebih mudah dibaca dan dipahami.

Analysis Questions

Analysis questions sangat penting dalam proses analisis data. Pertanyaan-pertanyaan ini dapat membantu mengarahkan analisa kita, memberikan tujuan yang jelas, dan memungkinkan mengidentifikasi informasi yang paling relevan. Berikut adalah beberapa pertanyaan analisis yang dibuat:

1. **Berapa Total Penyewaan Sepeda?**
   Menghitung jumlah total penyewaan sepeda (count) sebagai KPI utama. Ini memberikan gambaran umum tentang seberapa banyak sepeda yang disewakan.

2. **Bagaimana Performance pengguna casual dan pengguna terdaftar?**
   - Penyewaan Pengguna Casual: menghitung rata-rata pengguna sepeda casual.
   - Penyewaan Pengguna Terdaftar: menghitung rata-rata pengguna sepeda terdaftar.
   - Perbandingan Antara Pengguna Casual dan Terdaftar: membandingkan rata-rata pengguna sepeda casual dan terdaftar.
3. **Bagaimana Perfromance penyewaan sepeda berdasarkan waktunya?**
   - Penyewaan Sepeda per Musim: Menghitung rata-rata penyewaan sepeda untuk setiap musim.
   - Penyewaan Sepeda per Bulan: Menghitung rata-rata penyewaan sepeda dalam sebulan.
   - Penyewaan Sepeda per Minggu: Menghitung rata-rata penyewaan sepeda dalam seminggu.
   - Penyewaan Sepeda per Hari: Menghitung rata-rata penyewaan sepeda untuk dalam sehari.
   - Penyewaan Sepeda per Jam: Menghitung rata-rata penyewaan sepeda untuk setiap jam.
   - Penyewaan Sepeda per Hari Kerja: Menghitung rata-rata penyewaan sepeda pada hari kerja.

- Penyewaan Sepeda per Hari Libur: Menghitung rata-rata penyewaan sepeda pada hari libur (weekend).

4. **Bagaimana Total Penyewaan Sepeda Berdasarkan Cuaca?**
   Menghitung total penyewaan sepeda untuk setiap jenis cuaca. Ini dapat membantu memahami bagaimana cuaca mempengaruhi penyewaan sepeda.

5. **Bagaimana Pengaruh Suhu, Kelembaban, dan Kecepatan Angin Terhadap Penyewaan Sepeda?**
   Menghitung total penyewaan sepeda berdasarkan suhu, kelembaban, dan kecepatan angin. Hal ini dapat membantu memahami bagaimana pengaruhnya terhadap penyewaan sepeda.

Analyze Process

Proses analisis yang dilakukan menggunakan SQL untuk menjawab pertanyaan-pertanyaan analisis yang sudah dibuat sebelumnya. Berikut adalah analisisnya:

1. **Berapa Total Penyewaan Sepeda?**

```
SELECT SUM(cnt) FROM rental_bike_sharing
```

Results | Messages

	(No column name)
1	3292679

2. **Bagaimana Performance pengguna casual dan pengguna terdaftar?**
   - Penyewaan Pengguna Casual: menghitung total pengguna sepeda casual.

```
SELECT SUM(casual) FROM rental_bike_sharing
```

Results | Messages

	(No column name)
1	620017

- Penyewaan Pengguna terdaftar: menghitung total pengguna sepeda terdaftar.

```
SELECT SUM(registered) FROM rental_bike_sharing
```

Results | Messages

	(No column name)
1	2672662

- Perbandingan Antara Pengguna Casual dan Terdaftar: membandingkan rata-rata pengguna sepeda casual dan terdaftar.

```
SELECT SUM(registered) AS total_registered,
       SUM(casual) AS total_casual,
       SUM(registered) - SUM(casual) AS different_value
FROM rental_bike_sharing
```

Results | Messages

	total_registered	total_casual	different_value
1	2672662	620017	2052645

3. **Bagaimana Perfromance penyewaan sepeda berdasarkan waktunya?**

- Penyewaan Sepeda per Musim: Menghitung rata-rata penyewaan sepeda untuk setiap musim.

```
SELECT season,
       SUM(count) as total_rental
FROM rental_bike_sharing
GROUP BY season
```

Results | Messages

	season	total_rental
1	winter	471348
2	summer	1061129
3	spring	918589
4	fall	841613

- Penyewaan Sepeda per Bulan: Menghitung rata-rata penyewaan sepeda dalam sebulan.

```
SELECT MONTH(date) AS month_name,
       SUM(count) AS total_rental
FROM rental_bike_sharing
GROUP BY MONTH(date)
ORDER by MONTH(date)
```

Results Messages

	month_name	total_rental
1	1	134933
2	2	151352
3	3	228920
4	4	269094
5	5	331686

- Penyewaan Sepeda per Minggu: Menghitung rata-rata penyewaan sepeda dalam seminggu.

```
SELECT DATEPART(wk, date) AS week,
SUM(count) AS total_rental
FROM rental_bike_sharing
GROUP BY DATEPART(wk, date)
ORDER BY total_rental DESC
```

Results Messages

	week	total_rental
1	23	83887
2	39	83396
3	38	82385
4	33	81516
5	43	81286
6	28	80663

- Penyewaan Sepeda per Hari: Menghitung rata-rata penyewaan sepeda untuk dalam sehari.

```
SELECT DAY(date) AS day,
       SUM(count) AS total_rental
FROM rental_bike_sharing
GROUP BY DAY(date)
ORDER BY total_rental DESC
```

Results Messages

	day	total_rental
1	17	118255
2	15	115677
3	20	114697
4	14	112406
5	4	112335
6	10	111645

- Penyewaan Sepeda per Jam: Menghitung rata-rata penyewaan sepeda untuk dalam setiap jam.

```
SELECT hour,
       SUM(count) AS total_rental
FROM rental_bike_sharing
GROUP BY hour
ORDER BY total_rental DESC
```

Results Messages

	hour	total_rental
1	17	336860
2	18	309772
3	8	261001
4	16	227748
5	19	226789
6	13	184919

- Penyewaan Sepeda per Hari Kerja: Menghitung rata-rata penyewaan sepeda pada hari kerja (weekday).

```
SELECT SUM(count) as total_rental
FROM rental_bike_sharing
WHERE weekday NOT IN ('saturday', 'sunday')
```

Results Messages

	total_rental
1	2370845

- Penyewaan Sepeda per Hari Kerja: Menghitung rata-rata penyewaan sepeda pada hari libur (weekend).

```
SELECT SUM(count) as total_rental
FROM rental_bike_sharing
WHERE weekday IN ('saturday', 'sunday')
```

Results Messages

	total_rental
1	921834

4. Bagaimana Total Penyewaan Sepeda Berdasarkan Cuaca?

```
SELECT weathersit,
       SUM(count) as total_rental
FROM rental_bike_sharing
GROUP BY weathersit
ORDER BY total_rental DESC
```

Results Messages

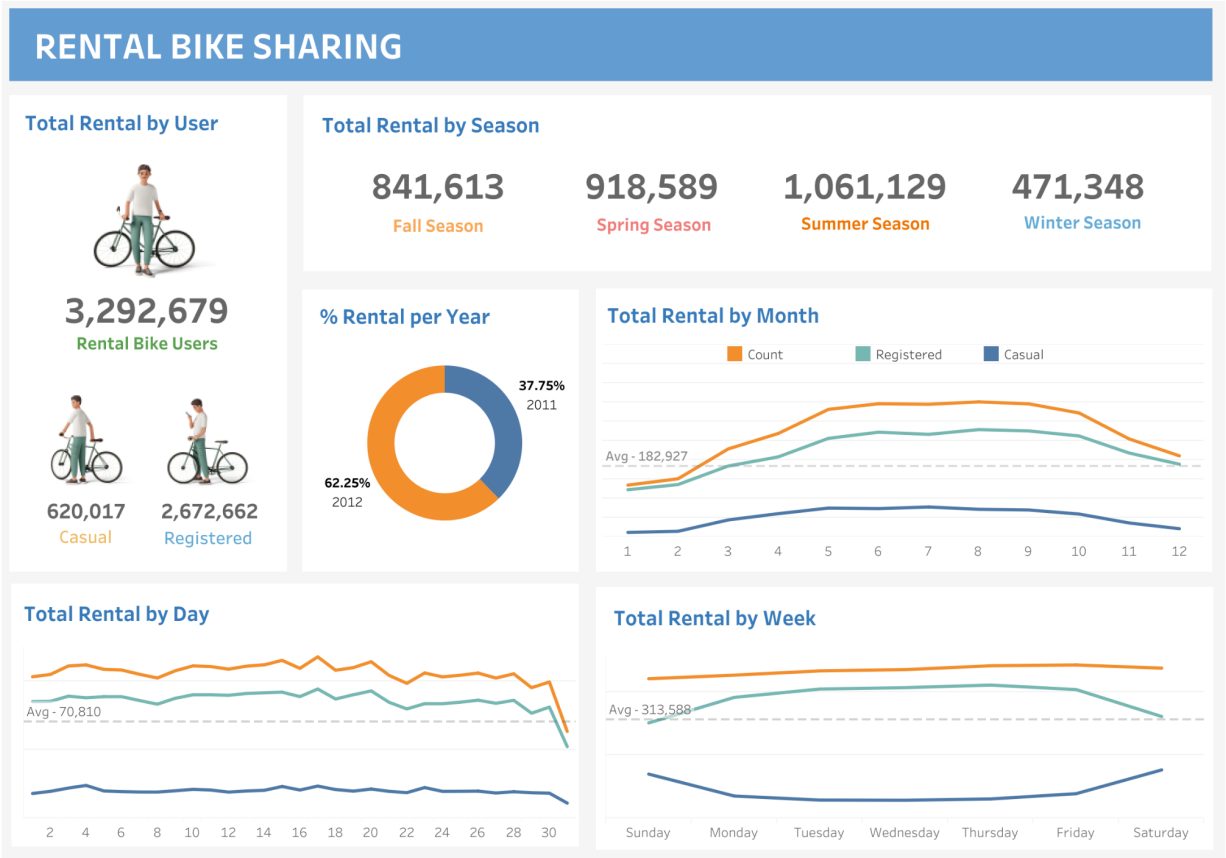
	weathersit	total_rental
1	1	2338173
2	2	795952
3	3	158331
4	4	223

5. **Bagaimana Pengaruh Suhu, Kelembaban, dan Kecepatan Angin Terhadap Penyewaan Sepeda?**

   Variabel suhu, kelembaban, dan kecepatan angin berbanding lurus dan saling mempengeharui pada variabel musim, misalnya jika musim kemarau maka suhu, kelembaban, dan kecepatan angin umumnya akan menunjukkan angka yang sama pada di musim kemarau. Oleh sebab itu variabel ini tidak kita analysis dan diwakilkan oleh variabel musim dan cuaca saja.

Visualisasi Data

Setalah melakukan proses analisis, selanjutnya adalah membuat dashboard untuk lebih memudahkan bisnis dalam melihat insight yang ada di dalam data, melihat tren rental sepeda, dan memudahkan untuk mengambil keputusan. Saya membuat dashboard menggunakan Tableau Public, hasilny adalah seperti ini:



Kesimpulan

Dari proses persiapan data, analisis data dan visualisasi data, didapatkan beberapa kesimpulan yaitu:

1. Total pengguna rental sepeda adalah sebanyak 3.292.679 yang terdiri dari 620.017 pengguna casual dan 2.672.662 pengguna terdaftar.
2. Penyewaan sepeda paling banyak pada musim panas dengan total 1.061.129, lalu diikuti dengan musim semi, musim gugur, dan musim dingin.
3. Dalam 2 tahun yaitu tahun 2011 dan 2012, penyewaan paling banyak pada tahun 2012.
4. Penyewaan paling banyak berdasarkan bulan terjadi antara bulan Mei sampai September, dikarekan pada bulan-bulan ini merupakan musim panas.
5. Total rental per hari relatif sama, tapi total tertingginya berada pada hari ke-17 dan menuju akhir bulan penyewaan menurun pada di hari ke-30 menurun drastis.
6. Berdasarkan hari dalam seminggu terdapat pola yang menarik dimana total penyewa sepeda dilihat stabil, tapi berdasarkan pengguna yang terdaftar hari weekend sediki yang menyewa sepeda, sedangkan pengguna casual di hari weekend lebih banyak menyewa sepeda.