Case Study Cyclistic Bike Share

A. The Cyclistic

Pada tahun 2016, Cyclistic meluncurkan layanan berbagi sepeda yang sukses. Sejak saat itu, program ini telah berkembang menjadi armada sebanyak 5.824 sepeda yang dilengkapi pelacakan lokasi (*geotracked*) dan dapat dikunci di jaringan yang terdiri dari 692 stasiun di seluruh Chicago. Sepeda-sepeda ini bisa dibuka dari satu stasiun dan dikembalikan ke stasiun mana pun dalam sistem kapan saja.

Hingga saat ini, strategi pemasaran Cyclistic bergantung pada peningkatan kesadaran umum dan pendekatan terhadap segmen konsumen yang luas. Salah satu pendekatan yang membantu mencapai hal ini adalah fleksibilitas dalam paket harga yang ditawarkan: tiket perjalanan tunggal (*single-ride passes*), tiket harian penuh (*full-day passes*), dan langganan tahunan (*annual memberships*). Pelanggan yang membeli tiket perjalanan tunggal atau harian penuh disebut sebagai pengguna kasual, sedangkan pelanggan yang membeli langganan tahunan disebut anggota Cyclistic.

B. Tujuan Study Case

Menganalisis perbedaan perilaku antara pengguna kasual dan anggota tahunan layanan berbagi sepeda Cyclistic di Chicago. Wawasan dari analisis ini akan digunakan untuk merancang strategi pemasaran yang mendorong pengguna kasual agar beralih menjadi anggota tahunan.

1. Ask

1.1 Busines Task

- Bagaimana anggota tahunan dan pengguna kasual menggunakan sepeda Cyclistic secara berbeda?
- Mengapa pengguna kasual akan membeli keanggotaan tahunan Cyclistic?
- Bagaimana Cyclistic dapat menggunakan media digital untuk memengaruhi pengguna kasual agar menjadi anggota?

1.2 Stakeholders

- Lily Moreno (Marketing Director).
- Tim Data Analyst.
- Tim Eksekutif Cyclistic.

2. Prepare

2.1 Data Use

Dataset yang digunakan dalam study kasus ini adalah data historis perjalanan sepeda Cyclistic yang tersedia di Rstudio atau jika tidak mau menggunakan Rstudio bisa diunduh https://divvy-tripdata.s3.amazonaws.com/index.html gunakan dataset Divvy 2019 Q1 dan Divvy 2020 Q1.

2.2 Data Save

- **Name File:** Cyclistic_TripData_2019Q1_2020Q1_Analysis.xlsx

2.3 Information about our dataset

Dataset ini merupakan gabungan data perjalanan sepeda yang dikumpulkan dari sistem penyewaan sepeda Cyclistic selama kuartal pertama tahun 2019 dan 2020. Data perjalanan ini mencakup informasi lengkap mengenai setiap perjalanan, termasuk waktu mulai dan selesai, lokasi stasiun awal dan akhir, jenis sepeda yang digunakan, serta tipe pengguna (anggota tahunan atau pengguna kasual).

Dataset berisi sekitar 791.956 perjalanan dengan detail yang memungkinkan analisis perilaku pengguna sepeda dalam berbagai konteks waktu dan lokasi. Data 2019 menggunakan format waktu dan stasiun yang sudah lama, sehingga sudah distandarisasi dan digabung dengan data 2020 di excel workbook untuk analisis yang konsisten.

2.4 Structure Data

Kolom	Keterangan
ride_id	ID unik untuk setiap perjalanan dimana tahun 2019 bertype numeric dan 2020 bertipe string, namun ditahun 2019 valuenya ditambah 2019_ (2019_21742443) agar bisa dibedakan datanya dengan 2020 karena formatnya sudah dijadikan 1 dan distandarisasi dan bertype data string.
rideable_type	Jenis sepeda yang digunakan dan hanya ada 1 type sepeda yaitu docked_bike
started_at	Tanggal dan waktu perjalanan dimulai
ended_at	Tanggal dan waktu perjalanan selesai
start_station_name	Nama stasiun pemberangkatan
start_station_id	ID stasiun pemberangkatan
end_station_name	Nama stasiun tujuan
end_station_id	ID stasiun tujuan
start_lat	Latitude lokasi stasiun pemberangkatan (pada data 2019 data tidak tersedia)
start_lng	Longitude lokasi stasiun pemberangkatan (pada data 2019 data tidak tersedia)
end_lat	Latitude lokasi stasiun tujuan (pada data 2019 data tidak tersedia)
end_lng	Longitude lokasi stasiun tujuan (pada data 2019 data tidak tersedia)
member_casual	Tipe pengguna: member (anggota tahunan) atau casual (pengguna kasual/tiket harian)

2.5 Accessibility and privacy of data

Dataset ini bersifat publik dan disediakan oleh *Motivate International Inc.*, nama dan label dalam dataset telah dimodifikasi agar sesuai dengan perusahaan fiktif Cyclistic**.** Dalam hal privasi**,** dataset ini tidak mengandung informasi identitas pribadi. Semua data pengguna telah dianonimkan sepenuhnya. Akses ke dataset ini tidak memerlukan izin khusus dan dapat diunduh secara bebas untuk analisis data non-komersial.

2.6 ROCCC Data

- **Reliable**: Sumber data terpercaya dan berasal dari sistem berbagi sepeda nyata.
- **Original**: Data mentah dari sistem pelacakan sepeda.
- **Comprehensive**: Meliputi informasi perjalanan lengkap selama dua kuartal berbeda.
- **Current**: Mewakili waktu terkini saat data tersedia (Q1 2019 & Q1 2020).
- **Cited**: Diberikan oleh penyedia layanan asli dan digunakan dalam konteks studi yang sah.

3. Process

3.1 Tool

Saya akan memfokuskan analisis menggunakan **Python** karena jumlah data yang digunakan cukup besar (791.956 baris), sehingga membutuhkan alat yang fleksibel dan efisien untuk melakukan pembersihan (data cleaning) dan analisis data. Python juga mendukung berbagai pustaka visualisasi yang memadai untuk mendukung pengambilan keputusan serta memudahkan proses penyampaian hasil analisis kepada para pemangku kepentingan.

3.2 Data Cleaning

No	Tanggal	Langkah	Deskripsi	Kode / Penanganan	Hasil / Tujuan
1	22 Juni 2025	Menggabungkan dua dataset Divvy_Trips_2019_Q1 dengan Divvy_Trips_2020_Q1	Digabungkan dengan excel dengan cara file csv diubah kedalam bentuk tabel menggunakan text to column lalu dijadikan satu tabel	Text_to_column	Mendapat satu dataset lengkap Cyclistic_TripData_2019Q1_2020Q1_Analyze untuk diload ke Python
Data Cleaning					
2		Load dataset	Import dataset ke dalam python menggunakan pd.read_csv	`pd.read_csv`	Mendapatkan satu data set df_all_trip
3		Melihat jumlah column & baris	-	`df_all_trip.shape`	Terdapat 791956 rows dan 13 column
4		Melihat informasi dataset	-	`df_all_trip.info()`	
5		Konversi Kolom Tanggal	Mengubah kolom started_at, ended_at ke format datetime.	`pd.to_datetime(df['started_at'])` `pd.to_datetime(df_all_trip['ended_at'])`	Memungkinkan pengolahan waktu seperti durasi dan hari.
6		Konversi end_station_id	Mengubah type data end_station_id dari float ke int	`df_all_trip['end_station_id'].fillna(-1).astype(int)`	Agar data bisa seragam dengan start_station_id dan sesuai format
7		Cek type data	-	`df_all_trip.dtypes`	Untuk melihat apakah ada type data yang perlu diubah
8		Rename nama column	Rename nama columns 'Started_at', 'start_lng', 'end_lng'	`df_all_trip.rename(columns={nama_column:nama_change_column},` `inplace=True)`	Agar sesuai standart dan tampilannya bagus

9		Feature Engineering	Feature Engineering (ride_length) untuk melihat durasi perjalanan pemakain sepeda dan day_of_week	`df_all_trip['ride_length'] =`
				`df_all_trip['ended_at'] -`
				`df_all_trip['started_at'`
					Untuk digunakan dalam proses analysist

				] df_all_trip['day_of_week'] = df_all_trip['started_at']. dt.weekday	
10		Mengubah posisi columns	Menempatkan ride_length dan day_of_week setelah ended_at	df_all_trip = df_all_trip[columns]	Agar tampilan lebih rapi
11		Backup Data	-	df_all_trip.copy(deep=True)	Untuk proses validasi
12		Cek missing value	Mengecek semua nilai yang hilang dalam columns maupun baris	df_all_trip.isnull().sum()	Untuk melihat data yang kosong agar tidak terjadi kesalahan dalam analisis data
13		Cek missing value di end_station_name	Karena sebelumnya pada pengecekan missing value seluruh columns ada 1 nilai NaN di end_station_name kita akan mencari tahu dan menggantinya	df_all_trip[df_all_trip['end_station_name'].isna()]	Agar formatnnya lebih konsisten
14		Mengganti missing value pada end_station_name	Setelah di cari tahu bahwa nilai end_station_name yang nan memiliki nilai ended_at < start_at dimana kebanyakan terjadi pada start_station_name dan end_station_name di 'HQ RQ' saya memutuskan akan mengisi dengan 'HQ RQ'	df_all_trip['end_station_name'].fillna('HQ RQ')	Tidak ada lagi nilai NaN pada end_station_name
15		Mencari & mengganti nilai end_station_id = -1	Karena pada proses penggantian type data pada end_station_id dari float ke	df_all_trip[df_all_trip['end_station_id'] == -1] df_all_trip['end_station_id'].replace(-1, 675)	Agar formatnya lebih konsisten dan sesuai dengan station_id real life

			int terdapat nilai NaN maka nilainya akan diberikan -1. Dan sekarang akan diganti valuenya dengan 675 karena merupakan station_id dari 'HQ RQ'		
16		Cek data duplicat	Cek duplicat semua kolom agar tidak ada data yang sama yang mengganggu analisis data dan hasil yang tidak akurat	`df_all_trip.duplicated().sum()`	Tidak ada data duplicat
17	24 Juni 2025	Cek unique value	Melakukan pengecekan unique value pada column ride_id, rideable_type, day_of_week, member_casual.	`df_all_trip[col].unique()`	Untuk mengecek apakah ada value diluar kategori atau ada kategori yang sama cuman salah penulisan
18		Cek konsistensi value	Cek konsistensi value start_station_name dan start_station_id serta end_station_name dan end_station_id	`start_station_check = df_all_trip.groupby('start_station_name')['start_station_id'].nunique()` `start_station_duplicates = start_station_check[start_station_check > 1]`	Menemukan station_name yang memiliki lebih dari satu station_id
19		Cek anomali	Cek anomali ended_at < started_at	`df_all_trip[df_all_trip['ended_at'] < df_all_trip['started_at']]`	Untuk mendeteksi outlier karena tidak mungkin ended_at < started_at
20		Cek anomali	Cek anomali ended_at = started_at, karena artinya sepeda ini tidak bergerak sama sekali dan tidak sesuai dengan tujuan analisis di awal atau memang error sistem	`df_all_trip[df_all_trip['started_at'] == df_all_trip['ended_at']]`	Untuk mendeteksi outlier dan fokus pada tujuan analisis data

21		Filter duration trip	Menghapus data ended_at < started_at dan ended_at = started_at dengan melakukan filtering	`df_all_trip[df_all_trip['ended_at'] > df_all_trip['started_at']]`	Agar data konsisten sesuai dengan tujuan analisis dan menghapus outlier yang akan mengakibatkan hasil yang tidak akurat
22		Cek spasi berlebih	Melakukan pengecekan terhadap spasi berlebih pada columns bertype data string (ride_id, rideable_type, start_station_name, end_station_name, member_casual)	`df_all_trip[col].astype(str).str.strip()`	Untuk konsistensi data, dan tidak terjadi kesalahan pada analisis data dan hasil yang tidak akurat
23		Menghitung panjang data	Mengecek panjang ride_id dengan fungsi len	`df_all_trip['ride_id'].str.len()`	Ternyata panjang data tidak konsisten pada data ditahun 2019 semua panjangnya sama yaitu 13 karakter sedangkan pada tahun 2020 panjang datanya ada yang 9, 10, 11, 12, 16 mungkin karena kesalahan sistem atau fromat berubah, karena tidak berpengaruh pada proses analisis data saya biarkan saja
24		Standarisasi value	Standarisasi value pada column member_casual agar huruf pertama memakai huruf besar dengan fungsi capitalize	`df_all_trip['member_casual'].str.capitalize()`	Untuk format data yang lebih konsisten
25		Standarisasi value	Mengganti value pada column rideably_type dari docked_bike menjadi Docked	`df_all_trip['rideable_type'].replace({'docked_bike': 'Docked Bike'})`	Agar tampilan data lebih konsisten

			Bike		
26		Convert value	Menjadikan ride_length dari type timedelta menjadi minute	`df_all_trip['ride_length'].dt.total_seconds() / 60`	Untuk memudahkan dalam proses perhitungan

27		Filter data	Filter data ride_length < 1, karena data bisa menjadi outlier, tidak ada perjalanan sepeda < 1 menit hal ini mungkin disebabkan karena aplikasi yang erro, kegagalan sistem, atau memang sepeda tidak jadi dipakai	`df_all_trip[df_all_trip['ride_length'] < 1]`
				
				`df_all_trip[df_all_trip['ride_length'] >= 1]`
					Menghilangkan outlier

28		Standarisasi ride_length	Karena ride_length bertype data float yang mungkin memiliki banyak angkat dibelakan koma maka kita hanya akan menampilkan dua angka dibelakang coma dengan fungsi round	`df_all_trip['ride_length'].round(2)`	Untuk konsistensi data
29		Drop column ride_id_length	Menghapus ride_id_length karena sudah tidak digunakan	`df_all_trip.drop(columns=['ride_id_length'])`	Untuk tampilan yang lebih rapi
30		Standarisasi value start_station_name dan end_station_name	Pada saat melakukan pengecekan data saya melihat terdapat (*) pada statuion_name yang pada kenyataannya di real life tidak ada jadi saya akan mencari dan menghapusnya	`df_all_trip['start_station_name'].str.replace(r'\(\*\)', '', regex=True).str.strip()`	Untuk tampilan yang lebih rapi
31	25 Juni 2025	Mengisi nilai NaN di columns start_lat, start_long, end_lat, end_long pada tahun	Melakukan pengisian nilai NaN pada columns start_lat, start_long, end_lat, end_long pada tahun 2019 dengan data	`df_all_trip['start_lat'] = df_all_trip['start_lat'].fillna(df_all_trip['start_lat_filled'])`	Untuk memastikan lokasi stasiun tetap akurat meskipun data awal tidak lengkap.

		2019 dengan data 2020	start_lat, start_long, end_lat, end_long 2020 berdasarkan station_id prosesnya yaitu : 1. Take All Coordinates From Start Station 2. Take All Coordinates From ENnd Station 3. Change The Name Columns to be Mergeable 4. Combining and Cleaning Station Coordinates Data 5. Filling NaN Values in Start Coordinates via Station ID Mapping 6. Filling NaN Values in End Coordinates via Station ID Mapping 7. Delete Temporary Columns	`df_all_trip['end_lat'] = df_all_trip['end_lat'].fillna(df_all_trip['end_lat_filled'])`	
32		Identifikasi column start_lat, start_long, end_lat, end_long yang masih NaN	Melihat apakah masih ada data yang NaN ditahun 2019	`df_all_trip[df_all_trip[['start_lat', 'start_long', 'end_lat', 'end_long']].isna().any(axis=1)]`	Deteksi nilai NaN yang station_id mungkin tidak cocok dengan tahun 2020
33		Identifikasi station_id 2019 yang tidak ada ditahun 2020	Melihat station_id 2019 yang tidak cocok dengan station_id 2020	`missing_2019[~missing_2019['start_station_id'].isin(station_2020_ids)][['start_station_id', 'start_station_name']].drop_duplicates()`	Untuk menemukan apa penyebab data masih NaN
34		Mengisi column start_lat, start_long, end_lat, end_long yang masih NaN	Ada 6 station_id ditahun 2019 yang datanya tidak ada ditahun 2020 hal tersebut menjadikan 4 columns titik	`df_all_trip['start_lat'].fillna(df_all_trip['start_station_id'].map(station_coords_lat))`	Mengisi nilai yang masih NaN agar data menjadi lengkap

			koordinat tersebut masih NaN pada proses mining data, untuk itu station_id yang tidak cocok dicari secara manual melalui website resmi Divvy Chicago dengan bantuan Chat GPT lalu melakukan pengimputan data secara manual	`df_all_trip['start_long'].fillna(df_all_trip['start_station_id'].map(station_coords_long))`	
35		Standarisasi start_lat, start_long, end_lat, end_long value	Setelah dicek rentang datanya melalui fungsi describe ternyata value dari start_lat, start_long, end_lat, end_long tidak sesuai dengan titik ordinate di real life saya menemukan nilai max dari masing-masing kolom sangat tinggi yaitu kisaran 381.586865 untuk latitude dan 810.684242 untuk longitude yang dimana nilai ini sangat jauh dari titik ordinat real life di chichago dimana latitude maksimum 41.97823 dan latitude minimum 41.69055 untuk longitude maksimum -87.55208 dan longitude minimum -87.80320 hal ini dikarenakan salah bacah titik atau koma pada saat proses membaca csv	`applymap(lambda x: x / 10 if abs(x) > 100 else x`	Untuk standarisasi data

36		Cek rentang data	Cek rentang data untuk semua columns	`df_all_trip.describe()`	Untuk melihat apakah rentang data sudah sesuai dengan real life data

37		Deteksi data abnormal pada ride_length	Setelah proses pengecekan rentang data ditemukan nilai maksimum dari ride_length yaitu 177200.370000 atau sekitar 123.06 hari dan angka ini tidak sesuai dengan durasi perjalanan yang digunakan oleh pengguna sepeda biasanya, untuk itu data ini akan dihapus agar fokus pada tujuan analysis data	`df_all_trip[df_all_trip['ride_length'] > 1440]`
				
				`df_all_trip[df_all_trip['ride_length'] <= 1440]`
					Deteksi outlier dan menghapusnya

38		Drop columns station_name_x station_name_y	Menghapus columns station_name_x station_name_y karena sudah tidak digunakan lagi	`df_all_trip.drop(['station_name_x','station_name_y'], axis=1)`	Untuk dataset lebih rapi
39		Final dataset	Melihat final dataset yang sudah selesai dibersihkan dengan menggunakan fungsi style pada Python Pandas agar tampilannya lebih menarik	`styled.set_properties(subset=pd.IndexSlice[:, :], **{'text-align': 'center'}, axis=0)`	Untuk melihat final dataset dengan tampilan yang rapi

3.3 Data Validation

No	Item Validasi	Kolom Terkait	Deskripsi / Tujuan	Status	Catatan
1	Count of columns & rows	All dataset	Untuk melihat jumlah final columns dan rows	Done	Rows final = 783803, dan columns final = 15
2	Tipe data	All columns	ride_id = object rideable_type = object started_at = datetime64[ns]	Done	Sudah dikonfirmasi dan sesuai dengan type data

			ended_at = datetime64[ns] ride_length = float64 day_of_week = int64 start_station_name = object start_station_id = int64 end_station_name = object end_station_id = int64 start_lat = float64 start_long = float64 end_lat = float64 end_long = float64 member_casual = object		
3	Missing value	All Columns	Tidak boleh ada NaN	Done	Semua kolom tidak ada NaN. NaN pada end_station_name diisi dengan 'HQ RQ' dan NaN di titik koordinate 2019 diisi dengan titik koordinate 2020 berdasarkan station_id dan station_id 2019 yang tidak cocok dengan 2020 dicari dan diisi secara manual
4	Duplicate	All rows	Tidak boleh ada duplikat data	Done	Sudah difilter dan ride_id diverifikasi unik
5	Nilai unik	rideable_type, member_casual, day_of_week	rideable_type = Docked Bike, member_casual = 'Member' atau 'Casual', day_of_week = 1, 2, 3, 4, 5, 6, 7	Done	Sudah distandarisasi
6	Rentang data	ride_length, start_lat, start_long, end_lat, dan end_long	ride_length dikonversi ke menit dan harus maksimum 1 menit dan maksimal < 24 jam dan start_lat, start_long, end_lat, dan end_long harus sesuai dengan titik ordinat real life di chichago dimana latitude maksimum 41.97823 dan latitude minimum 41.69055 untuk longitude maksimum -	Done	Sudah sesuai dengan rentang data real life

			87.55208 dan longitude minimum -87.80320		
7	Pengisian NaN untuk titik koordinat 2019 dengan titik koordinat 2020	start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_long, end_lat, dan end_long	Apakah proses melakukan maining data sudah berhasil artinya proses pengisian titik koordinat 2019 sudah cocok antara station_id 2019 dengan station_id 2020	Done	Semua sudah benar sesuai station_id
8	Filter data	started_at & ended_at	ended_at <= started_at	Done	Semua value sudah ended_at > started_at dan baris yang tidak valid sudah dihapus
9	Filter data	ride_length	ride_length >= 1 menit & ride_length < 1440 menit	Done	Values sudah sesuai dengan rentang data dan outlier 123 hari sudah dihapus

4. Interpretasi Hasil Analisis

4.1 Statistic descriptive of ride_length & day_of_week

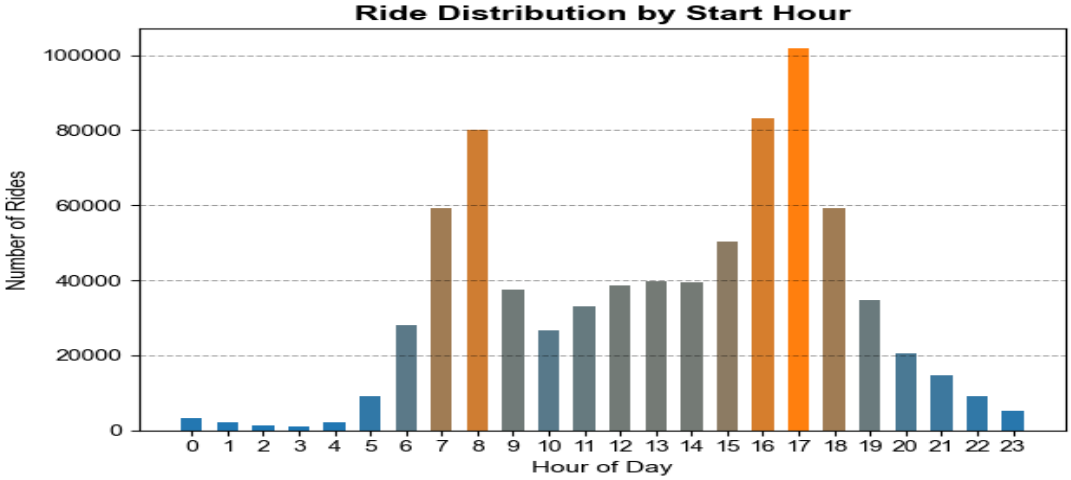
Statistik	ride_length	day_of_week
Count	780,382	780,382
Mean	12.42	3.99
Std	11.48	1.81
Min	1.00	1.00
25%	5.53	3.00
50%	8.98	4.00
75%	15.12	5.00
Max	119.98	7.00

Ride_length: total perjalanan setelah filter < 120 menit yaitu 780.382. dengan rata-rata durasi 12.42 menit dan separuh perjalanan berdurasi 8.98 menit. Bersepeda tersingkat selama 1 menit dan terpanjang 2 jam.variasi bersepeda cukup besar yaitu 11.48, artinya menunjukkan banyak pengguna sepeda adalah member. Namun dilihat dari standar deviasi dan nilai maximum yang tinggi menunjukkan ada pengguna casual yang melakukan perjalanan panjang.

Day_of_week: rata-rata bersepeda di hari mendekati Selasa/Rabu, hari Rabu adalah titik tengah frequensi dan variasi distribusi antar hari cukup merata. Maksimal bersepeda di hari sabtu dan minimal di hari minggu. Pola ini mendukung kebanyakan pengguna member bersepeda pada hari kerja dan casual bersepeda di weekend.

Kesimpulan: untuk memperkuat hasil analisis deskriptif pada ride_length dan day_of_week kita harus menganalisis ini antar member dan membandingkan hasilnya.

4.2 Count of Rides per hour



1. **Waktu Puncak Penggunaan (Peak Hours)**
   Jam paling sibuk untuk memulai perjalanan 08:00, 16:00, dan 17:00 jam ini konsisten dengan jam commuting kantor atau sekolah, artinya kebanyakan persepeda mayoritas pengguna memanfaatkan sepeda sebagai moda transportasi utama untuk bekerja atau kuliah.
2. **Pola Konsisten**
   Sementara itu, jam 11.00 hingga 14.00 menunjukkan volume penggunaan stabil, yang mengindikasikan adanya aktivitas santai, rekreasi, atau penggunaan non-commuting.
3. **Pengguna Sepeda pada Malam Hari Rendah**
   Penggunaan sepeda sangat rendah antara pukul 0:00 – 5:00 pagi, yang wajar karena ini adalah jam tidur atau sepi aktivitas kota.

4.3 Total ride_length per day_of_week

Day of Week	Total Ride Length
1 (Minggu)	1,325,520.22
2 (Senin)	1,307,891.75
3 (Selasa)	1,553,905.61
4 (Rabu)	1,522,103.38
5 (Kamis)	1,505,421.75
6 (Jum'at)	1,400,646.28
7 (Sabtu)	1,075,530.74

Total durasi perjalanan tertinggi terjadi pada hari kerja, terutama Selasa hingga Kamis, yang menunjukkan bahwa pesepeda sangat aktif dalam menggunakan sepeda untuk kebutuhan commuting. Sebaliknya, akhir pekan memiliki total durasi perjalanan yang lebih rendah. Untuk memperkuat hasilnya kita bisa melakukan analisis untuk member dan casual. Sehingga kita menemukan adanya peluang untuk mempromosikan casual menjadi member.

4.4 Statistic descriptive of ride_length by member_casual

Tipe Pengguna	Count	Mean	Std	Min	25%	Median (50%)	75%	Max
Casual	64,779	28.51	22.47	1.02	12.72	22.33	35.80	119.98
Member	715,603	10.96	8.52	1.00	5.33	8.50	13.73	119.92

Casual: memiliki jumlah perjalanan yang sedikit dengan rata-rata durasi perjalanan 28.51 menit. Dimana durasi terlama yaitu 119.98 menit dan durasi tersingkat yaitu 1.02 menit.

Member: memiliki jumlah perjalanan jauh lebih banyak dibandingkan casual, dengan rata-rata perjalanan 10.96 menit. Dimana durasi perjalanan terlama 119.92 menit dan tersingkat 1 menit.

Kesimpulan: Casual melakukan perjalanan 3 kali lipat jauh lebih lama dan bervariasi, sendangkan member cenderung melakukan perjalanan lebih pendek dan sering, yang menunjukkan pola pengguna yang lebih konsisten dan teratur.

4.5 Average and Median ride_length by member_casual and day_of_week

Day of Week	Casual Mean	Casual Median	Member Mean	Member Median
1 (Minggu)	32.60	25.90	12.50	9.18
2 (Senin)	24.55	19.15	10.71	8.42
3 (Selasa)	25.23	18.95	10.83	8.52
4 (Rabu)	29.03	22.49	10.83	8.47
5 (Kamis)	25.27	19.50	10.71	8.40
6 (Jum'at)	25.23	19.18	10.56	8.23
7 (Sabtu)	29.65	23.40	11.75	8.78

Casual: menunjukkan rata-rata durasi perjalanan yang tinggi di akhir pekan yaitu Sabtu 23.40 menit dan Minggu 25.90 menit.

Member: menunjukkan rata-rata durasi perjalanan yang kosisten setiap harinya yaitu (±8-9 menit).

Kesimpulan: casual lebih aktif diakhir pekan mungkin digunakan untuk aktivitas rekreasi atau aktivitas santai lainnya sedangkan member mengandalkan sepeda untuk perjalanan yang rutin harian mungkin untuk pergi kekantor.

4.6 Average and Median day_of_week by member_casual

Tipe Pengguna	Mean Day of Week	Median Day of Week
Casual	3.84	4.0
Member	4.01	4.0

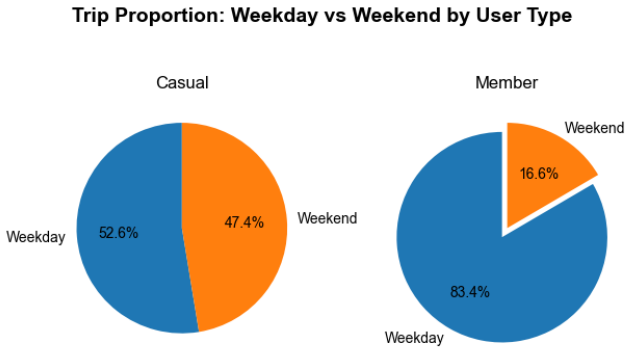
Casual: memiliki rata-rata sedikit lebih rendah daripada member yaitu 3.84 dan memiliki median yang sama dengan casual yaitu 4.
Member: memiliki rata-rata 4.01 dan median 4.
Kesimpulan: baik casual dan member paling sering bersepeda pada hari rabu. Namun casual cenderung mulai lebih aktif bersepeda dihari Selasa hingga Rabu, sedangkan member konsisten di hari Rabu dan hal ini kemungkinan konsisten untuk communting.

4.7 Proportion of Commute vs Non-Commute Rides by member_casual

Tipe Pengguna	Non-Commute Hours	Commute Hours
Casual	72.73%	27.27%
Member	50.37%	49.63%

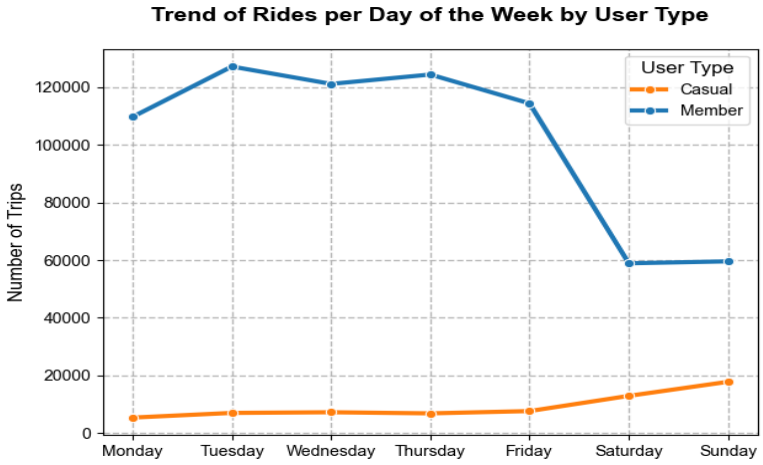
Casual: kebanyakan casual bersepeda di jam non commuting, hanya 27% pengguna casual memulai bersepeda di jam commuting, ini menjadi penguat analisis sebelumnya bahwa casual cenderung menggunakan sepeda untuk rekreasi atau aktivitas non_rutin.
Member: seimbang mengawali bersepeda di jam commuting dan non-commuting yang artinya member konsisten bersepeda untuk mobilitas harian.
Kesimpulan: casual menggunakan sepeda untuk rekreasional sedangkan member untuk rutinitas harian.

4.8 Proportion of Weekend vs. Weekday Rides by member_casual



Casual: lebih seimbang melakukan aktivitas bersepeda yaitu weekday 52.6% dan weekend 47.4%.
Member: sangat dominan menggunakan sepeda di weekday yaitu 83.4% dan weekende 16.6%.
Kesimpulan: walaupun seimbang dalam bersepeda antara weekday dan weekend namun hampir 50% pengguna casual bersepeda di akhir pekan ini semakin menguatkan bahwa casual menggunakan sepeda untuk aktivitas santai atau sekedar jalan-jalan. Sedangkan member sangat dominat bersepeda di weekday yang mencerminkan penggunaan sepeda lebih fungsional.

4.9 Count of Rides per day_of_week by casual_member



Casual: puncak bersepeda pengguna casual yaitu pada hari Sabtu = 12.913 dan Minggu = 17.797.

Member: melakukan perjalanan secara konsisten disepanjang hari kerja Senin – Jum'at.

Kesimpulan: ini semakin menguatkan analisis sebelumnya mengenai proporsi weekend dan weekday bahwa casual bersepeda pada weekend dan member bersepeda di weekday.

4.10 Top 5 Start Stations by member_casual

ID	Tipe Pengguna	Nama Stasiun	Jumlah Perjalanan (trip_count)	Tipe Stasiun
528	Casual	Streeter Dr & Grand Ave	2,659	Start Station
317	Casual	Lake Shore Dr & Monroe St	2,654	Start Station
468	Casual	Shedd Aquarium	1,809	Start Station
391	Casual	Millennium Park	1,341	Start Station
385	Casual	Michigan Ave & Oak St	987	Start Station
687	Member	Canal St & Adams St	13,747	Start Station
754	Member	Clinton St & Washington Blvd	13,378	Start Station
750	Member	Clinton St & Madison St	12,819	Start Station
912	Member	Kingsbury St & Kinzie St	8,677	Start Station
756	Member	Columbus Dr & Randolph St	8,461	Start Station

Casual: top lima station pengguna casual untuk memulai aktivitas bersepeda berada di lokasi wisata dan rekreasi utama Chicago. Ini menguatkan bahwa pengguna casua lbersepeda untuk tujuan rekreasi karena kemungkinan besar pengguna casual bersepeda di area sekitar dan kembali ke tempat yang sama dekat.

Member: top lima station pengguna casual untuk memulai aktivitas bersepeda berada di area pusat bisnis dan perkantoran yang semakin menguatkan bahwa pengguna member bersepeda untuk pergi bekerja atau kebutuhan harian lainnya.

Kesimpulan: Dari lima besar start stasiun, terlihat pola yang konsisten. Pengguna casual cenderung menggunakan sepeda untuk rekreasi dan beraktivitas di sekitar lokasi wisata seperti Streeter Dr & Grand Ave, Millennium Park, dan Shedd Aquarium. Sementara itu, member memanfaatkan sepeda untuk commuting, dengan titik awal dan akhir yang konsisten di pusat transit dan perkantoran seperti Canal St dan Clinton St.

4.11 ride_length Statistics by member_casual and Season

Tipe Pengguna	Musim	Mean	Median	Std Dev	Jumlah (Count)
Casual	Spring	30.76	24.50	23.07	38,799
Casual	Winter	25.15	18.97	21.10	25,980

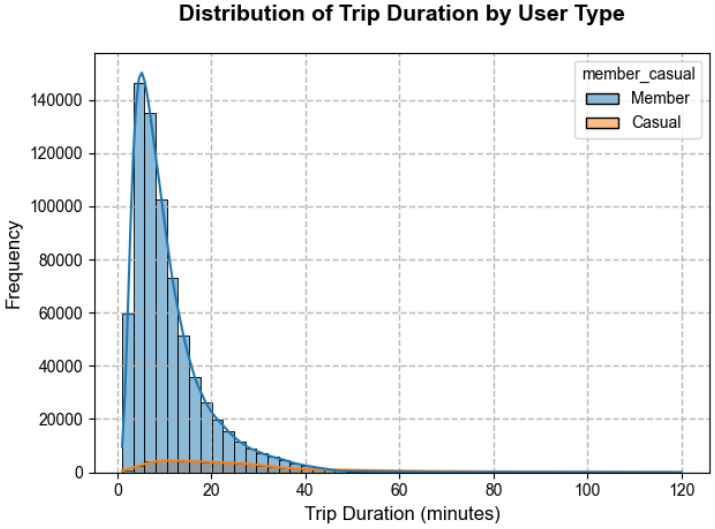
Member	Spring	11.66	8.95	9.14	263,776
Member	Winter	10.56	8.25	8.10	451,827

Casual: memiliki rata-rata durasi bersepeda 30.76 menit dimusim Spring dan 25.15 di Winter dengan variasi perjalanan 23.07 menit di Spring dan 21.10 di Winter.

Member: memiliki rata-rata durasi bersepeda 11.66 menit dimusim Spring dan 10.56 di Winter dengan variasi perjalanan 9.14 menit di Spring dan 8.10 di Winter.

Kesimpulan: Pengguna casual cenderung untuk perjalanan yang lebih panjang dan bervariasi, dengan aktivitas puncak di musim semi. Member menggunakan sepeda lebih sering tapi dengan perjalanan yang lebih singkat dan konsisten, menandakan pemakaian untuk keperluan commuting. Variasi tinggi pada durasi casual menunjukkan beragam tujuan (rekreasi, hiburan, olahraga) dibandingkan member yang cenderung rutin. Jumlah perjalanan member jauh lebih banyak, jadi meskipun durasi per perjalanan lebih pendek, total penggunaan mereka sangat besar.

4.12 Distribution of ride_length by member_casual



1. **Dominasi Durasi Pendek**
   Mayoritas perjalanan berdurasi pendek**,** terutama antara 0–20 menit**.** Ini terlihat dari puncak grafik (mode) yang sangat tinggi di kisaran waktu ini. Durasi pendek ini umum dalam penggunaan untuk commuting cepat atau perjalanan antar lokasi dekat.
2. **Perbedaan Pola antara Member dan Casual**
   Member: Memiliki frekuensi yang jauh lebih tinggi di hampir semua durasi, terutama pada durasi pendek. Menunjukkan bahwa anggota aktif menggunakan layanan secara rutin, sering kali untuk perjalanan harian seperti ke kantor atau ke transportasi umum.
   Casual: Jumlah perjalanan jauh lebih sedikit**,** tetapi cenderung merata sedikit lebih panjang pada ekor distribusi. Hal ini mengindikasikan penggunaan yang lebih santai atau rekreasional**,** bukan commuting rutin.
3. **Distribusi Eksponensial Turun**
   Semakin lama durasi perjalanan, semakin sedikit frekuensinya, baik untuk member maupun casual. Hal ini wajar karena perjalanan panjang bisa lebih melelahkan atau tidak efisien dibanding moda transportasi lain.

Kesimpulan: member cenderung menggunakan sepeda lebih frekuen dan efisien**.** Mendominasi perjalanan dengan durasi singkat, menandakan pola commuting atau penggunaan rutin. Casual Menggunakan sepeda lebih jarang. Cenderung melakukan

perjalanan dengan durasi sedikit lebih panjang, mengarah pada penggunaan rekreasional atau eksplorasi kota.

5. Kesimpulan akhir

Tujuan dalam analisis ini yaitu untuk menganalisis perbedaan perilaku antara pengguna casual dan member dalam bersepeda. Wawasan dari analisis ini akan digunakan untuk merancang strategi pemasaran yang mendorong pengguna casual agar menjadi member. Dataset yang digunakan yaitu data peseda quartal pertama tahun 2019 dan 2020.

5.1 Perbedaan member & casual dalam bersepeda

Aspek	Member (Tahunan)	Casual (Non-member)
Frekuensi	Melakukan lebih banyak perjalanan (frequensi tinggi) dengan rata-rata perjalanan 10.96 menit.	Frequensi lebih rendah dalam melakukan perjalanan bersepeda tetapi cenderung lebih merata artinya bersepeda untuk hal yang santai. Rata-rata durasi bersepeda 28.51 menit.
Durasi	Durasi perjalanan yang lebih konsisten dengan median 8.50 menit. Durasi tersingkat 1 menit dan terlama 119.92 menit	Durasi terlama yaitu 119.98 menit dan durasi tersingkat yaitu 1.02 menit.
Hari Penggunaan	83.4% konsisten aktif di weekday, terutama dihari Selasa – Kamis.	Lebih seimbang melakukan aktivitas bersepeda yaitu weekday 52.6% dan weekend 47.4%. yang artinya hampir 50% perjalanan dominan di weekend.
Jam Penggunaan	Jam sibuk saat bersepeda di jam commuting: 7–9 pagi & 4–6 sore.	Di luar jam commuting, siang atau sore santai. Hanya 27% pengguna casual memulai bersepeda di jam commuting.
Stasiun Favorit	5 start station paling favorit yaitu semuanya berada diarea perkantoran dan pusat bisnis.	5 start station paling favorit berada didekat tempat wisata dan area rekreasi.
Musiman	Konsisten selama musim Spring dan Winter.	Pengguna casual cenderung untuk perjalanan yang lebih panjang dan bervariasi, dengan aktivitas puncak di musim semi.

5.2 Mengapa pengguna casual akan menjadi member di Cyclistic?

Berdasarkan hasil analisis perbandingan antara casual dan member. Casual berpotensi menjadi member karena:

- Efisiensi biaya, walaupun memiliki frequensi yang rendah dibandingkan member tapi casual memiliki durasi bersepeda yang konsisten dan rata-rata durasi dalam bersepeda sangat tinggi dibandingkan member yaitu 28.51 menit. Jika ini dilakukan secara rutin setiap hari maka total biaya akan menjadi lebih tinggi dibandingkan member.
- Casual memiliki frequensi yang sangat tinggi di weekend namun begitu proporsi yang dilakukan di weekday tetap lebih tinggi daripada weekend yaitu 52.6%, ini membuka peluang untuk mengarahkan casual untuk beraktivitas di weekday juga misalnya untuk pergi ke kantor.

- Kebutuhan rutin/fleksibel, jika casual 2–3 kali seminggu, maka langganan tahunan lebih menguntungkan.
- Kenyamanan dan fleksibilitas waktu: Dengan menjadi member, pengguna mendapat akses yang lebih mudah dan tidak perlu memikirkan perhitungan biaya setiap perjalanan.

Dengan menyoroti penghematan biaya, fleksibilitas, dan akses mudah**,** pengguna casual bisa diyakinkan bahwa menjadi member memberi nilai lebih besar jika mereka menggunakan layanan secara lebih reguler.

5.3 Bagaimana Cyclistic dapat menggunakan media digital untuk memengaruhi pengguna kasual agar menjadi anggota?

Berdasarkan perbedaan perilaku antara pengguna casual dan member, Cyclistic dapat menggunakan media digital secara strategis untuk mendorong konversi pengguna casual menjadi member dengan pendekatan yang *personalized*, relevan, dan berbasis data. Berikut strategi utamanya:

1. **Optimalisasi Waktu Penayangan Iklan Berdasarkan Pola Aktivitas Casual**
   Strategi:
   - Menayangkan iklan digital (Google Ads, Instagram, Facebook) secara intensif pada hari dan jam ketika pengguna casual paling aktif, yaitu menjelang akhir pekan (Jumat sore hingga Minggu pagi) serta di luar jam sibuk kerja (non-commuting hours seperti pukul 10:00–15:00 atau malam hari).
   - Mengirim email marketing setiap Kamis atau Jumat berisi rekomendasi rute santai dan event bersepeda akhir pekan.

   Alasan: Pengguna casual cenderung bersepeda untuk rekreasi dan lebih aktif di akhir pekan atau waktu senggang. Menyesuaikan jadwal penayangan iklan dengan pola ini meningkatkan peluang keterlibatan dan konversi.
2. **Komunikasi Manfaat Keanggotaan secara Visual dan Edukatif**
   Strategi:
   - Menggunakan media sosial dan email untuk menampilkan visual yang menggugah, seperti:
     "Suka bersepeda di akhir pekan? Nikmati lebih banyak dengan akses tanpa batas sebagai anggota Cyclistic!"
     "Lebih hemat, lebih fleksibel. Gabung sebagai member mulai akhir pekan ini!"
   - Membuat konten edukatif seperti: Perbandingan biaya antara pengguna casual dan anggota dan testimoni pengguna yang awalnya casual lalu menjadi member.

   Alasan: Banyak pengguna casual belum memahami keuntungan finansial dan fleksibilitas sebagai member. Konten edukatif dan emosional dapat membentuk persepsi positif dan mempercepat pengambilan keputusan untuk berlangganan.
3. **Targeting Geografis Berdasarkan Data Start Station Populer**
   Strategi: Menayangkan iklan digital dengan geotargeting di area sekitar lokasi wisata dan start station populer yang sering digunakan pengguna casual, seperti Streeter Dr & Grand Ave, dan Millennium Park.
   Alasan: Karena banyak pengguna casual memulai perjalanan dari lokasi rekreasi, penargetan geografis membuat iklan lebih relevan dan kontekstual, sehingga meningkatkan kemungkinan klik dan konversi.