

Google Data Analytics Case Study

How does a bike-share navigate speedy success?

2021-07-22

Scenario

Cyclistic is a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. We would like to understand how casual riders and annual members use Cyclistic bikes differently. Therefore, our team will design a new marketing strategy to convert casual riders into annual members.

Agenda

1. Business Task
2. Description of data used
3. Documentation of cleaning data
4. Analyze and Visualizations
5. A summary of your analysis
6. Top three recommendations based on the analysis

Business Task

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become memberships?

Data Preparation

We will use Cyclistic's historical trip data to analyze and identify trends. (Last one year)

- 2020 Q1
- 2019 Q4
- 2019 Q3
- 2019 Q2

The data has been made available by Motivate International Inc. under this [license](#).

Process data

1. Each csv has different column names, we need to fix these to be matched.
2. Added two columns for duration of ride and weekday (1-7).
3. For 2019 data, member and casual riders are described as "Subscriber" and "Customer". We will Change this to "member" and "casual" in R.
4. Needed to fix ride length and added some columns in R. (year, month, and age)

(Process 2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ride_id	rideable_t	started_at	ended_at	ride_length	day_of_week	start_stat	start_stat	end_stat	end_stat	start_lat	start_lng	end_lat	end_lng
2	EACB1913	docked_b	2020-01-21 20:06	2020-01-21 20:14	0:07:31		3 Western Ave	239 Clark St &	326	41.9665	-87.6884	41.9671	-87.6884	
3	8FED874C	docked_b	2020-01-30 14:22	2020-01-30 14:26	0:03:43		5 Clark St &	234 Southport	318	41.9616	-87.666	41.9542	-87.666	
4	789F3C21	docked_b	2020-01-09 19:29	2020-01-09 19:32	0:02:51		5 Broadway	296 Wilton Ave	117	41.9401	-87.6455	41.9402	-87.6455	
5	C9A388DA	docked_b	2020-01-06 16:17	2020-01-06 16:25	0:08:49		2 Clark St &	51 Fairbanks	24	41.8846	-87.6319	41.8918	-87.6319	
6	943BC3CB	docked_b	2020-01-30 8:37	2020-01-30 8:42	0:05:32		5 Clinton St	66 Wells St &	212	41.8856	-87.6418	41.8899	-87.6418	
7	6D9C8A69	docked_b	2020-01-10 12:33	2020-01-10 12:37	0:04:49		6 Wells St &	212 Desplaine	96	41.8899	-87.6343	41.8846	-87.6343	
8	31EB9B8F	docked_b	2020-01-10 13:07	2020-01-10 13:12	0:04:49		6 Desplaine	96 Wells St &	212	41.8846	-87.6446	41.8899	-87.6446	

(Process 3)

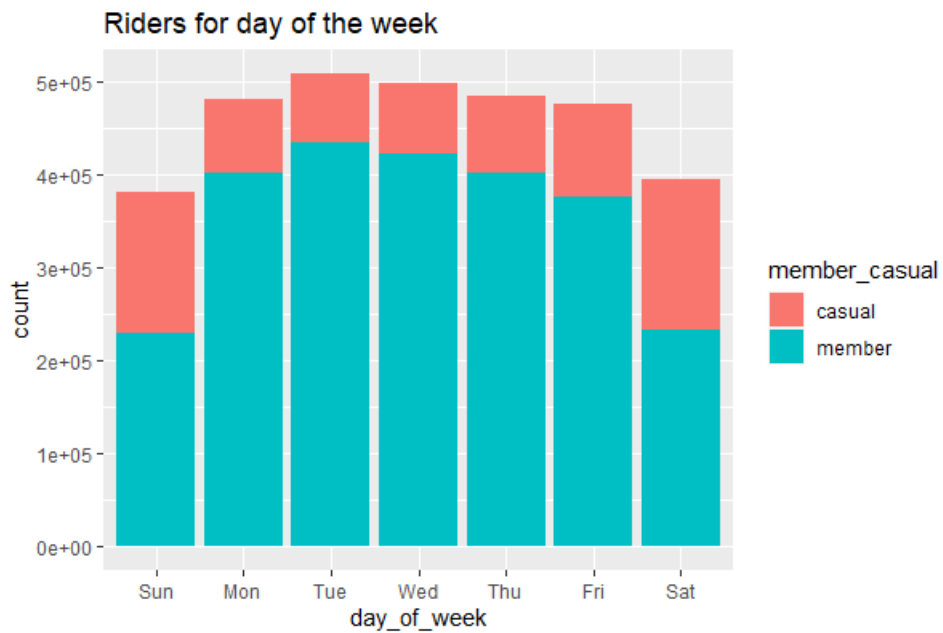
```
9 q1_2020 <- read.csv("R/bike_sharing/csv_file/Divvy_Trips_2020_Q1.csv")
10 q4_2019 <- read.csv("R/bike_sharing/csv_file/Divvy_Trips_2019_Q4.csv")
11 q3_2019 <- read.csv("R/bike_sharing/csv_file/Divvy_Trips_2019_Q3.csv")
12 q2_2019 <- read.csv("R/bike_sharing/csv_file/Divvy_Trips_2019_Q2.csv")
13
14 # Change value "Subscriber or Customer" to "member or casual" (2019 data)
15 q4_2019$member_casual <- gsub("Subscriber", "member", q4_2019$member_casual)
16 q4_2019$member_casual <- gsub("Customer", "casual", q4_2019$member_casual)
17
18 q3_2019$member_casual <- gsub("Subscriber", "member", q3_2019$member_casual)
19 q3_2019$member_casual <- gsub("Customer", "casual", q3_2019$member_casual)
20
21 q2_2019$member_casual <- gsub("Subscriber", "member", q2_2019$member_casual)
22 q2_2019$member_casual <- gsub("Customer", "casual", q2_2019$member_casual)
23
24
```

(Process 4)

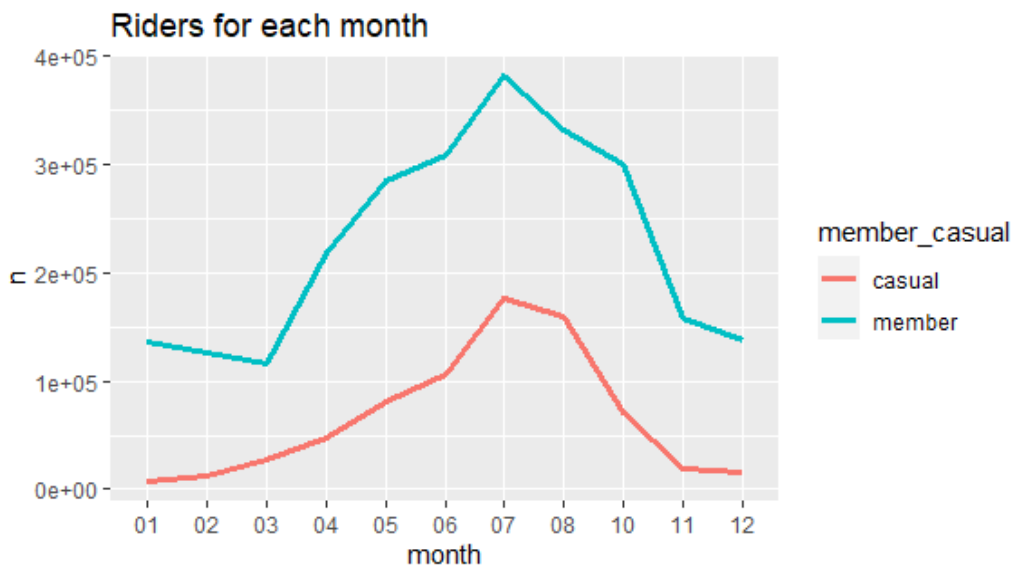
```
# Add columns
all_data$ride_length <- difftime(all_data$ended_at, all_data$started_at)/60
all_data$ride_length <- as.integer(all_data$ride_length)
all_data$month <- as.Date(all_data$started_at, format = "%Y-%m-%d")
all_data$month <- format(all_data$month, "%m")
all_data$year <- as.Date(all_data$started_at, format = "%Y-%m-%d")
all_data$year <- format(all_data$year, "%Y")
all_data$year <- as.integer(all_data$year)
all_data$age <- all_data$year - all_data$birthyear
```

Analyze

1. Number of riders for each weekday & month grouped by member and casual.



- Number of member riders is relatively high during weekday.
- Number of casual riders is high on weekend (Saturday and Sunday)



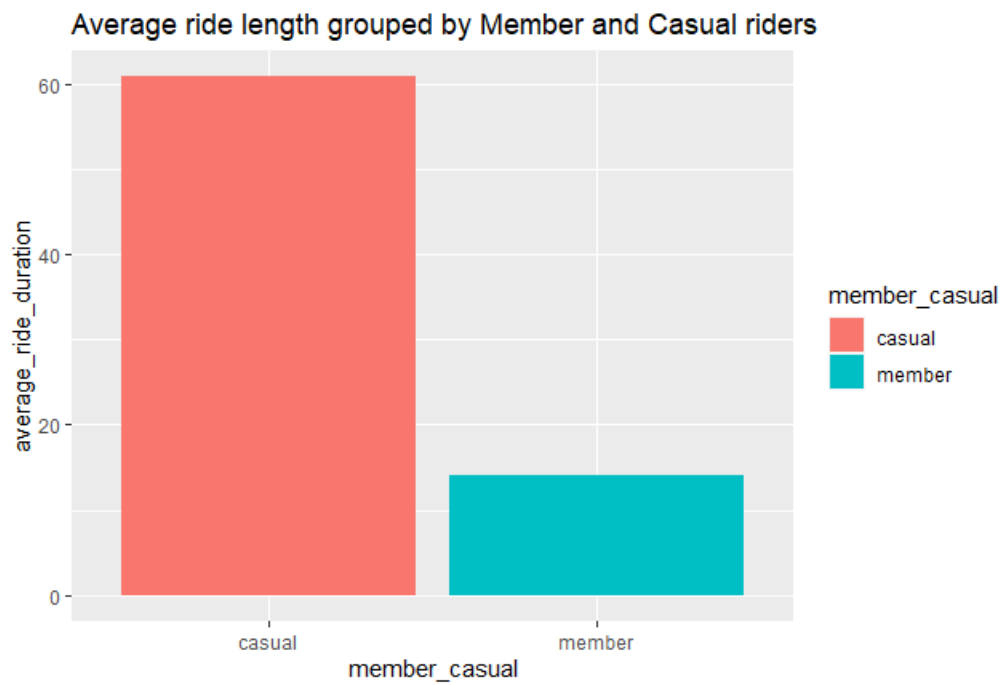
- Both number of riders are high during summer.
- Number of casual riders increase during summer vacation.

2. Ride duration for member and casual

Column “ride_length” is Character data type. In order to aggregate the data, we need to change the data type to numeric.

<Summary>

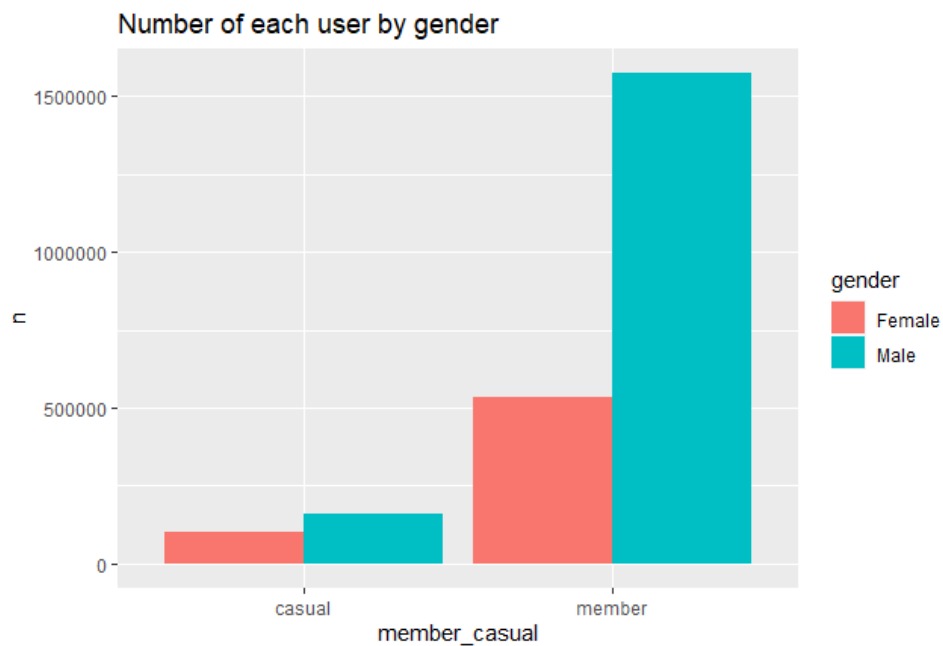
member_casual	average Ride duration
casual	60.8 minutes
member	14.1 minutes



Clearly, **Casual** riders ride bikes longer than Memberships.

3. Gender by member and casual riders

member_casual	gender	number
casual	Female	101820
casual	male	160652
member	Female	534752
member	male	1572865



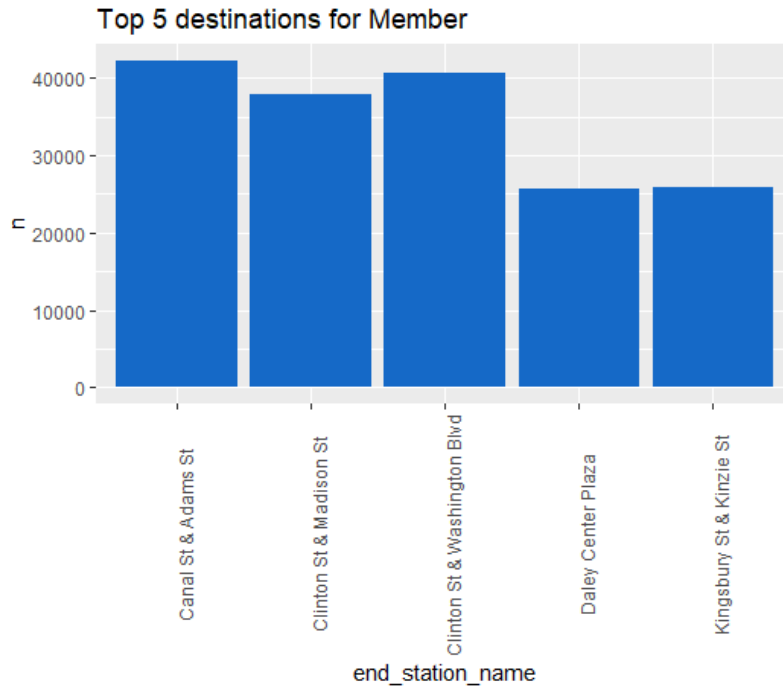
For the **Membership**, **Male** user has much higher number comparing to the Female user.

4. Investigate Destinations

We thought analyzing the user's destination can help us to understand more about the usage of the Cyclistic.

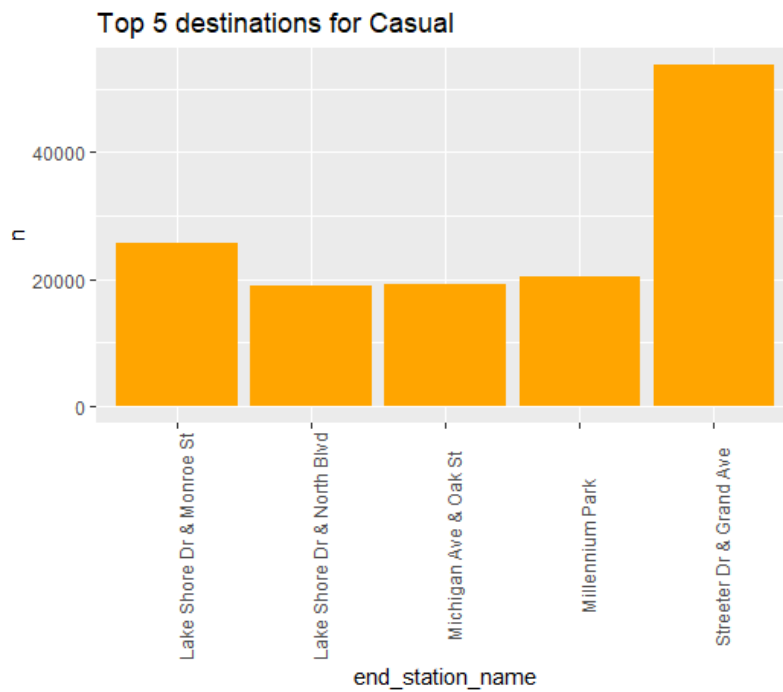
Top 5 Destinations for Member riders

end_station_name	member_casual	number
Canal St & Adams St	member	42280
Clinton St & Washington Blvd	member	40654
Clinton St & Madison St	member	37875
Kingsbury St & Kinzie St	member	25935
Daley Center Plaza	member	25729



Top 5 Destinations for Casual riders

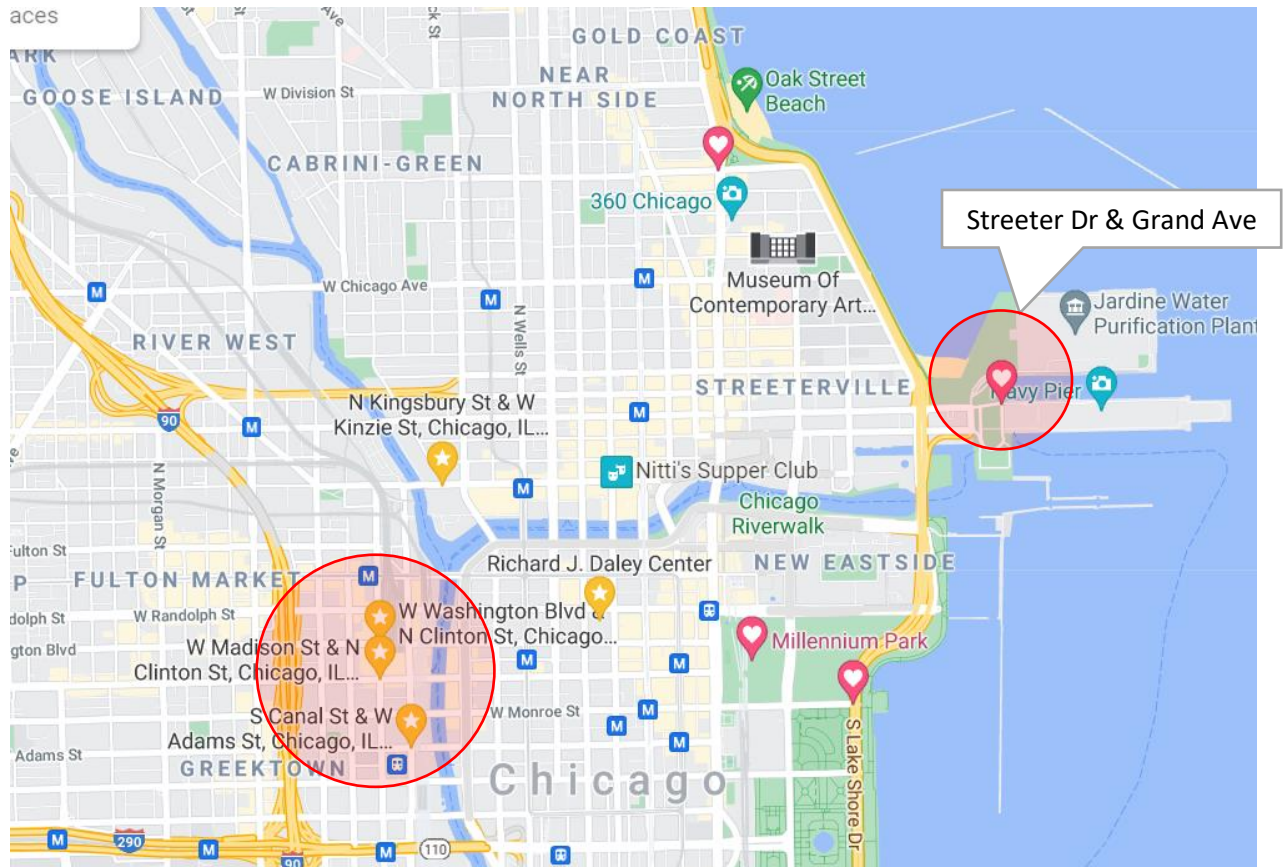
end_station_name	member_casual	number
Streeter Dr & Grand Ave	casual	53719
Lake Shore Dr & Monroe St	casual	25596
Millennium Park	casual	20266
Michigan Ave & Oak St	casual	19121
Lake Shore Dr & North Blvd	casual	19008



In the map below, I have marked Member's destination with Yellow Star and Casual's destination with Pink Heart.

★ = Member's destination

♥ = Casual's destination



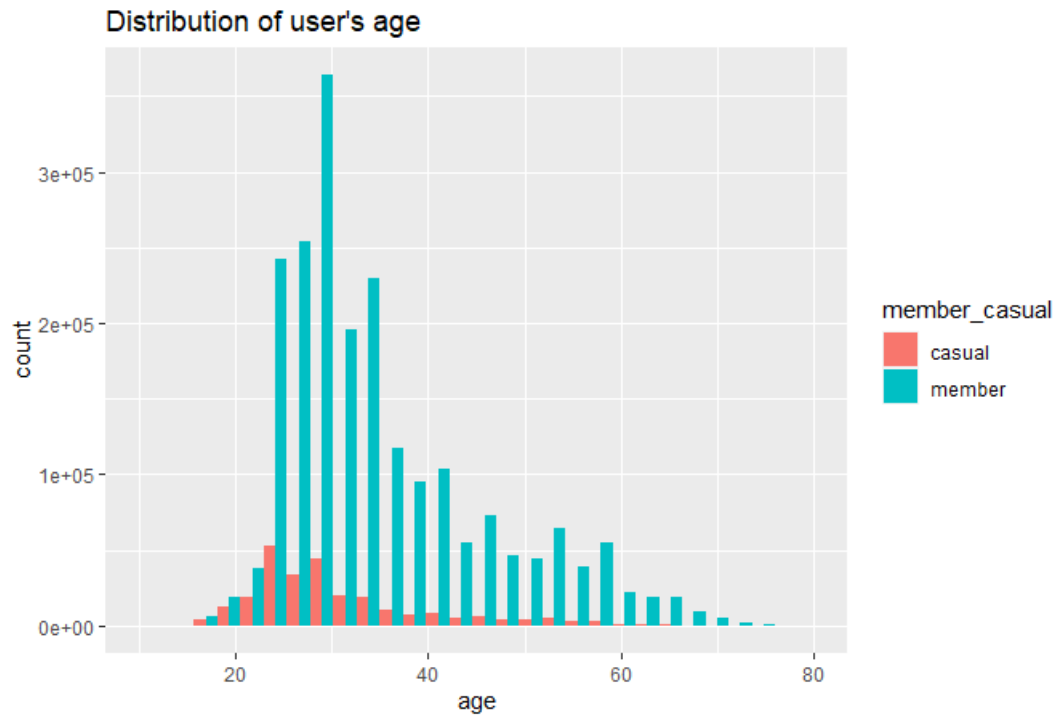
Member

- We think Member user's use Cyclistic for **commute**.
- Top 3 destinations are close to office buildings.

Casual

- We think Casual user's use Cyclistic for **leisure** purpose.
- Each destination is close to tourist spots such as park, beach and pier.
- Among the Top 5, "**Streeter Dr & Grand Ave**" has extremely high number.

5. Distribution of user's age grouped by member and casual riders



Member

- Relatively wide age range of people are member.
- Especially age between 25 – 35 is high.

Casual

- Compared to member users, casual users are younger.
- Age between 20 – 30 is high.

Summary of analysis

Member riders and Casual riders utilize our bikes quite differently.

Member

- Use Cyclistic for commute.
- A lot of users ride our bikes during weekdays (Mon-Fri).
- High number of male users.
- Wide range of age users (25-40).
- Stations are close to office buildings.

Casual

- Use Cyclistic for leisure.
- A lot of users ride our bikes during weekends (Sat-Sun).
- Not big difference between gender.
- Relatively younger age (25 – 30)
- Stations are close to tourist spots.

Top 3 Recommendations

1. Create an app for members which offers:
 - Information around the place (Tourism spots, Food...etc)
 - Discount (Tickets, Food, Drink...etc)
2. Invest on ads around the tourism spots. Focus on “Streeter Dr & Grand Ave”
3. Add different type of bikes to widen age group
 - Bikes with a safe child seat for family
 - Electric bikes for elderly people