Ask Step

Question: **How do annual members and casual riders use Cyclistic bikes differently**?

Guiding questions:

* What is the problem you are trying to solve?
  + How do annual members use Cyclistic bikes differently from the casual riders? (**actual problem**)
* How can your insights drive business decisions?
  + Knowing key factors differentiating annual members from casual riders can help guide business decision to encourage casual riders to convert to annual membership.

Key tasks:



1. Identify the business task
2. The **primary (general) business task is to maximize the number of annual memberships**
3. Our goal (**specific**) as the analytics team is to **design marketing strategies aimed at converting casual riders into annual members**
4. Ideal state would be to have all riders sign up for membership whereas the current state is a mix of casual and annual riders
5. Consider key stakeholders
6. Key stakeholders include Cyclistic executive team and Lily Moreno
7. They expect us to determine ways to **maximize the number of annual memberships**

The current business task is to compare patterns of casual riders from annual members in hopes of designing marketing strategies aimed at converting casual riders into annual members thus maximizing the number of annual memberships.

Prepare Step

Guiding questions:

* Where is your data located?
  + The data is located at <https://divvy-tripdata.s3.amazonaws.com/index.html>
* How is the data organized?
* Are there issues with bias or credibility in this data according to ROCCC?
  + Reliable: the data comes from the company directly so it is reliable
  + Original: the data comes from a first-party source so it is original
  + Comprehensive:
  + Current: the data spans the last 12 months so it is recent
  + Cited: the data is NOT cited
* How are you addressing licensing, privacy, security, and accessibility?
  + Rider ID, which is the **only** identifying factor differentiating riders, gives no personal information on the customer that used the bike
* How did you verify the data’s integrity?
* How does it help you answer your question?
* Are there any problems with the data?
  + There appears to be significant missing data on the starting/ending locations for multiple months which may hinder identifying distance as a factor in differentiating casual from annual members

Key tasks:

1. Download data and store it appropriately
2. Identify how it’s organized
   1. In an excel spreadsheet
3. Sort and filter the data

Process

Guiding questions:

* What tools are you choosing and why?
  + Microsoft excel was used to easily sort/filter missing data as well as adding additional columns
* Have you ensured your data’s integrity?
  + The data is missing several information on station names/id/lat/long as well as containing inaccuracies with starting/ending time. To ensure data integrity, I have removed aforementioned inaccuracies. The observations missing their latitude and longitude have been removed also to ensure credibility of the ride length remains accurate.
* What steps have you taken to ensure that your data is clean?
  + First, I remove rows with missing end lat/long as it is difficult to determine distance when either of the lat/long are missing 🡪 decreases credibility of the ride length
  + Second, I remove rows that have longer start time vs end time as that does not make sense
  + Third, I insert two more columns, “ride length” and “day of week” to determine the length of the ride as well as which day the customer used the bike. These data may be useful in determine the key differences between casual and annual members
  + Fourth, I round all lat/long values to the nearest second decimal to clean format
* How can you verify that your data is clean and ready to analyze?
  + One way to verify that the data is clean and ready to analyze is by following the process of verification.
    - Consider the business problem: “how can we maximize the number of annual members?”
    - Consider the project goal: “how can we convert casual riders to annual members?” / “what are the differences between casual riders and annual members?”
    - Based on the first two questions, is the data appropriate?
      * Using the data on most recent 12 months of rider history, we can use ride length and weekday used to determine whether annual members use the service longer than casual members and how the day of bicycle used differ between two groups. If bicycles are used more frequently on certain weekdays/weekend, designing a strategies marketing towards those days may be beneficial in converting casual riders to annual members.
  + Another way to verify is by ensuring the data is complete, accurate, up to date, and consistent without duplicate observations.
    - All inconsistent and missing data relevant to the business problems have been removed from the data (i.e., missing ending lat/long & time serviced used greater than time serviced finished)
* Have you document your cleaning process so you can review and share those results?

Key tasks:

1. Check the data for errors
2. Choose your tools
3. Transform the data so you can work with it effectively
4. Document the cleaning process

Analyze

Guiding questions:

* How should you organize your data to perform analysis on it?
* I prepared the data by typecasting started\_at/ended\_at as datetime values and converting ride\_duration in terms of hours. I also added a new column called distance to determine the distance the bike was used to travel. The important aspect of these steps was taken to allow for smooth transformation in plots and its labels.
* Has your data been properly formatted?
* The data has been formatted to ensure no NA are included and typecasting were done corresponding to the type of the fields (e.g., fields related to datetime being typecasted as such)
* What surprises did you discover in the data?
* What trends or relationship did you find the data?
* Casual riders tend to spend more time on the bicycles than annual members and they also tend to travel further than annual members. However, there are higher number of annual members using the service more than casual riders.
* How will these insights help answer your business questions?
* From this, we might gather that casual riders may be using the bike services only on certain occasions where they might need to travel greater distances. This corresponds to how casual riders spend more time on their bike than annual members. On the flipside, annual members may use the service more regularly (e.g., such as for commute purposes) as seen from the higher user count and lower distance/ride duration recorded. It would be useful to identify whether the annual members use the service more **frequently** by exploring duplicate rider ids or other indicators which may imply patterns of repetition (such as what you might expect for one’s commute).

Key tasks:

1. Aggregate your data so it’s useful and accessible
2. Organize and format your data
3. Perform calculations
4. Identify trends and relationships

Share

Compelling presentation requires considerations of the:

1. Characters
   1. people affected by your project (e.g., stakeholders)
   2. Lily Moreno (director of marketing and your manager) and the executive team (higher level team so focus the presentation to answer their questions directly without going into too much details. You can put additional details in the appendix)
2. Setting
   1. Describing the current situation
   2. The team is looking to increase the number of annual members by developing a marketing strategies to convert casual riders to annual members.
3. Plot
   1. Often describes the conflict (that are addressed/resolved in the project) such as challenge from a competitor, an inefficient process, or a new opportunities that should compel the characters to act
   2. My task was to determine key differences between casual riders and annual members to identify factors which could be used to convert casual riders to annual members.
4. Big reveal
   1. Showing how your data can solve the problem your characters are facing
   2. The preliminary trends based on the data show that casual riders often use the services to travel greater distances and for longer period of time compared to annual members. Furthermore, the top 5 destinations visited by the users all lie near an attraction (e.g., parks, docks, theatres) and the majority demographic consist of casual riders hinting a possibility that casual riders use the service for leisurely purposes. On the other hand, annual members may use the services more regularly as a mean to an end (e.g., commute). We also see that both demographics use the services more frequently during weekends than weekdays.
5. Aha moment
   1. Where you share your recommendation and why your decision will help the character
   2. We may be able to convert casual riders to annual members by promoting campaigns to incentivize casual riders to use the services more. This could include promoting eco-friendly agendas (e.g., encouraging uses of bikes as a commute than vehicles) or installing more stations near residential districts to allow for easier access

Guiding questions:

* Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?
* The preliminary trends based on the data show that casual riders often use the services to travel greater distances and for longer period of time compared to annual members. Furthermore, the top 5 destinations visited by the users all lie near an attraction (e.g., parks, docks, theatres) and the majority demographic consist of casual riders hinting a possibility that casual riders use the service for leisurely purposes. On the other hand, annual members may use the services more regularly as a mean to an end (e.g., commute). We also see that both demographics use the services more frequently during weekends than weekdays.
* What story does your data tell?
* The Cyclistic executive team wants to know how casual riders use the service differently than annual members in hopes of increasing annual membership. The current situation shows that casual riders use the service to travel further than annual members and for longer as well, however, annual members seem to use the services more than casual riders. The preliminary trends also showed that casual riders seemed to use the service predominately to travel to various attractions during weekends whereas annual members use the service more during weekdays. This may suggest that they are using the service as a means to an end such as commutes. Based on these findings/differences, we may be able to convert casual riders to annual members by promoting campaigns to incentivize casual riders to use the services more often rather than an occasional tool to travel longer distances. One such way could be to install more stations near residential districts to increase ease of access.
* How do your findings relate to your original question?
* The difference between casual vs annual is that casual riders using the services tend to travel further/longer than annual members and they seem to use them to travel to various attractions during weekends. The original question was “How do annual members use Cyclistic bikes differently from the casual riders?”
* Who is your audience? What is the best way to communicate with them?
* My audience are the executive teams and since they are higher-ups, the presentation should be directed at higher level discussions (i.e., broad ideas, getting directly into the conclusion without details)
* Can data visualization help you share your findings?
* Is your presentation accessible to your audience?

Act

Guiding questions:

* What is your final conclusion based on your analysis?
* Based on these findings/differences, we may be able to convert casual riders to annual members by promoting campaigns to incentivize casual riders to use the services more often rather than an occasional tool to travel longer distances. One such way could be to install more stations near residential districts to increase ease of access.
* How could your team and business apply your insights?
* What next steps would you or your stakeholders take based on your findings?
* Is there additional data you could use to expand on your findings?