

**GOOGLE DATA ANALYTICS CAPSTONE 1 - English and Francais / Étude de cas sur l'analyse de données Google.**



**Data Informations / Informations sur les données**

This [dataset](#) has been given by Google for use in case study 1 and has been made available under this [license](#). *Cet ensemble de données a été fourni par Google pour être utilisé dans l'étude de cas 1 et a été mis à disposition sous cette licence.* **Scenario** I am a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Lily Moreno, my manager, believes there is a very good chance to convert casual riders into members. I have a clear goal: Design marketing strategies aimed at converting casual riders into annual members.

*Je suis une analyste de données junior travaillant dans l'équipe d'analystes marketing de Cyclistic, une entreprise de vélos en libre-service à Chicago. Le directeur du marketing estime que le succès futur de l'entreprise dépend de la maximisation du nombre d'adhésions annuelles. Lily Moreno, ma manager, pense qu'il y a de très bonnes chances de convertir les coureurs occasionnels en membres annuels. J'ai un objectif clair : Concevoir des stratégies marketing visant à convertir les cyclistes occasionnels en membres annuels.*

**ASK / INTERROGER**

- **What is the problem you are trying to solve? / Quel est le problème que vous essayez de résoudre ?** Understanding the difference between casual riders and annuals member and how to convert casual riders into annual members. *Comprendre la différence entre les coureurs occasionnels et les membres annuels et comment convertir les coureurs occasionnels en membres annuels.* **Who are your stakeholders? / Qui sont vos stakeholders, managers ?** Cyclistic marketing analytics team, The director of marketing and my manager, Lily Moreno *Équipe d'analyse marketing Cyclistic, Le directeur du marketing et ma responsable, Lily Moreno.\**

**PREPARE / PRÉPARER**

- Are there issues with bias or credibility in this data? Does your data ROCCC? *Y a-t-il des problèmes de biais ou de crédibilité dans ces données ?* The data has no bias because there is no private informations about people (Card ID, Social Security Number, address etc..) *Les données n'ont aucun biais car il n'y a pas d'informations privées sur les personnes (ID de carte, numéro de sécurité sociale, adresse, etc.)*
- How did you verify the data's integrity? I verified the data's integrity by eliminate duplicates, checking unique values, trimming and by eliminate Na's. *J'ai vérifié l'intégrité des données en éliminant les doublons, en vérifiant les valeurs uniques et en éliminant les Na.*

**I will do this analysis in R (Rstudio) and I will also use Tableau for some charts. I will use data from 01/21 to 12/21 / Je ferai cette analyse avec R (Rstudio) et j'utiliserai également Tableau pour certains graphiques. J'utiliserai les données du 21/01 au 21/12**

Download the last 12 months of Cyclistic datasets in "casestudy" file, In Rstudio, install packages. Set working directory. / *Téléchargez les 12 derniers mois d'ensembles de données Cyclistic dans le fichier "casestudy", dans Rstudio, installez les packages. Définir le répertoire de travail.*

```
library(tidyverse)
install.packages("lubridate")
library("lubridate")
library(dplyr)
library(ggplot2)
```

Error in library(tidyverse): there is no package called 'tidyverse'  
Traceback:

```
1. library(tidyverse)
```

**Import datasets with new names into RStudio. / Importez des ensembles de données avec de nouveaux noms dans RStudio.**

```
jan <- read_csv("../input/tripdatacyclist/202101-divvy-tripdata.csv")
feb <- read_csv("../input/tripdatacyclist/202102-divvy-tripdata.csv")
mar <- read_csv("../input/tripdatacyclist/202103-divvy-tripdata.csv")
apr <- read_csv("../input/tripdatacyclist/202104-divvy-tripdata.csv")
may <- read_csv("../input/tripdatacyclist/202105-divvy-tripdata.csv")
june <- read_csv("../input/tripdatacyclist/202106-divvy-tripdata.csv")
july <- read_csv("../input/tripdatacyclist/202107-divvy-tripdata.csv")
aug <- read_csv("../input/tripdatacyclist/202108-divvy-tripdata.csv")
sept <- read_csv("../input/tripdatacyclist/202109-divvy-tripdata.csv")
oct <- read_csv("../input/tripdatacyclist/202110-divvy-tripdata.csv")
nov <- read_csv("../input/tripdatacyclist/202111-divvy-tripdata.csv")
dec <- read_csv("../input/tripdatacyclist/202112-divvy-tripdata.csv")
feb <- read_csv("../input/tripdatacyclist/202102-divvy-tripdata.csv")
mar <- read_csv("../input/tripdatacyclist/202103-divvy-tripdata.csv")
apr <- read_csv("../input/tripdatacyclist/202104-divvy-tripdata.csv")
may <- read_csv("../input/tripdatacyclist/202105-divvy-tripdata.csv")
june <- read_csv("../input/tripdatacyclist/202106-divvy-tripdata.csv")
```

```
july <-read_csv("../input/tripdatacyclist/202107-divvy-tripdata.csv")
aug <-read_csv("../input/tripdatacyclist/202108-divvy-tripdata.csv")
sept <-read_csv("../input/tripdatacyclist/202109-divvy-tripdata.csv")
oct <-read_csv("../input/tripdatacyclist/202110-divvy-tripdata.csv")
nov <-read_csv("../input/tripdatacyclist/202111-divvy-tripdata.csv")
dec <-read_csv("../input/tripdatacyclist/202112-divvy-tripdata.csv")
```

**Combine all files into a single dataframe.** / *Combinez tous les fichiers en une seule data frame.*

```
alltrips<- bind_rows(jan, feb, mar, apr, may, june, july, sept, nov,
oct, dec)
```

**PROCESS** / *Traiter*

**Check first parts of data frame.** / *Vérifiez les premières parties du bloc de data frame.*

```
head(alltrips)
str(alltrips)
summary(alltrips)
```

**Delete the columns that we don't need.** / *Supprimez les colonnes dont nous n'avons pas besoin.*

```
alltrips <- alltrips %>% select(-start_lat, -start_lng, -end_lat, -
end_lng, -end_station_id, -start_station_id)
```

**Check internal structure of data frame.** / *Vérifiez la structure interne de la data frame.*

```
str(alltrips)
```

**Add start date, month, day, year and day of weeks columns to data frame for more detailed analysis.** / *Ajoutez des colonnes de date de début, mois, jour, année et jour de la semaine au data frame pour une analyse plus détaillée.*

```
alltrips$date <- as.Date(alltrips$started_at)
alltrips$month <- format(as.Date(alltrips$date), "%m")
alltrips$day <- format(as.Date(alltrips$date), "%d")
alltrips$year <- format(as.Date(alltrips$date), "%Y")
alltrips$day_of_week <- format(as.Date(alltrips$date), "%A")
```

**View summary of data frame.** / *Afficher le résumé de la data frame.*

```
summary(alltrips)
```

**Add a column of ride durations** / *Ajouter une colonne de durées de trajet.*

```
alltrips$ride_length <-
difftime(alltrips$ended_at,alltrips$started_at)
```

**CLEAN DATA**

**Check and delete Na's** / *Vérifier et supprimer les Na*

```
any(is.na(alltrips))
sum(is.na(alltrips))
alltrips<- na.omit(alltrips)
sum(is.na(alltrips))
```

**Values of member\_casual column can not be different from member and casual.**

**Check if they are unique.** / Les valeurs de la colonne member\_casual ne peuvent pas être différentes de celles de member et casual. Vérifiez s'ils sont uniques. **Also values of rideable\_type column can not be different from classic\_bike, electric\_bike and docked\_bike. Check if they are unique.** / Les valeurs de la colonne rideable\_type ne peuvent pas être différentes de celles de classic\_bike, electric\_bike et docked\_bike. Vérifiez s'ils sont uniques.

```
unique(alltrips[c("member_casual")])
unique(alltrips[c("rideable_type")])
```

**It can be useful to create a season column. Format month column as numeric and create a season column. Check with view() and head()** / Il peut être utile de créer une colonne de saison. Formatez la colonne du mois en tant que numérique et créez une colonne de saison. Vérifier avec view() et head()

```
alltrips$month<-as.numeric(format(alltrips$date,"%m"))
alltrips <- alltrips %>%
  mutate(season =
    ifelse(month %in% c(12, 01, 02), "Winter",
           ifelse(month %in% c(03, 04, 05), "Spring",
                  ifelse(month %in% c(09, 10, 11), "Fall",
                         "Summer"))))
head(alltrips)
```

**it can also be helpful to know the hours they start riding. Add a starttime column.** / Il peut également être utile de connaître les heures auxquelles ils commencent à faire du vélo. Ajoutez une colonne starttime.

```
alltrips$starttime <- format(as.POSIXct(alltrips$started_at), format =
"%H:%M:%S")
```

**Quick check max and min of ride\_length.** / Vérification rapide sur max et min valeurs de la colonne ride\_length.

```
max(alltrips$ride_length)
min(alltrips$ride_length)
```

**Ride\_length can not be negative. Update data frame by applying a filter.** / Ride\_length ne peut pas être négatif. Mettez à jour la data frame en appliquant un filtre.

```
alltrips <- alltrips %>%
  filter(ride_length > 0)
```

**Check min time difference again.** / Verification de min ride\_length

```
min(alltrips$ride_length)
```

**Since I've finished cleaning the dataframe, I'm renaming it as cleantrips1.** / Depuis que j'ai fini de nettoyer la data frame, je la renomme en cleantrips1.

```
cleantrips1 <- alltrips
```

**ANALYSE Calculate average duration by member type.** / Calculez la durée moyenne par type de membre.

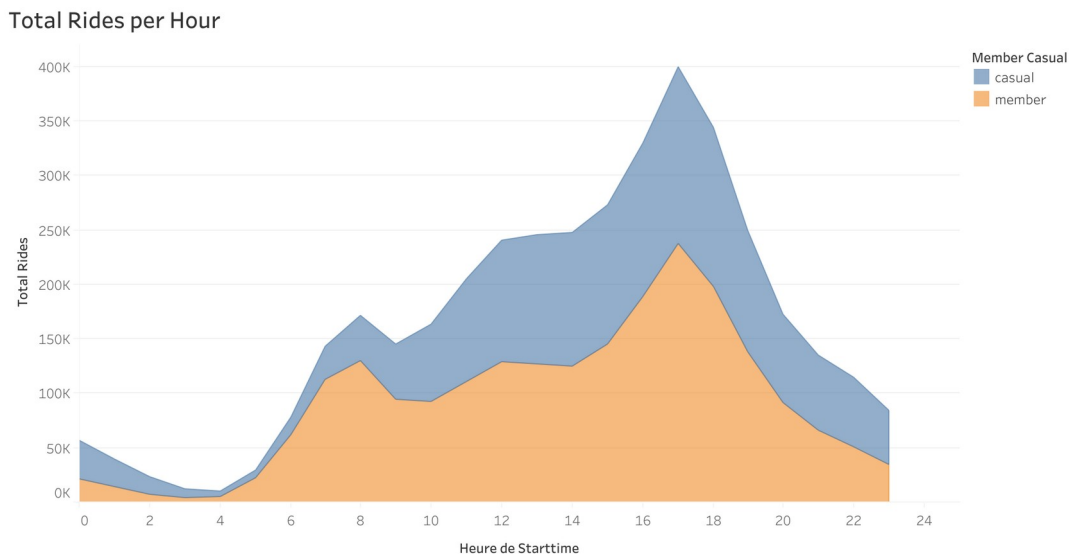
```
avgduration <- cleantrips1 %>% select(member_casual, ride_length) %>%  
  group_by(member_casual) %>%  
  summarize("Average Duration" = round(mean(ride_length), 0))
```

**Calculate ride\_length by seasons for each type of member.** / Calculez ride\_length par saisons pour chaque type de membre.

```
ggplot(cleantrips1, aes(season, ride_length, fill=member_casual)) +  
  geom_bar(stat='identity')
```

**Check total rides per hour (via Tableau).** / Calculez le nombre total de trajets par heure pour chaque type de membre.

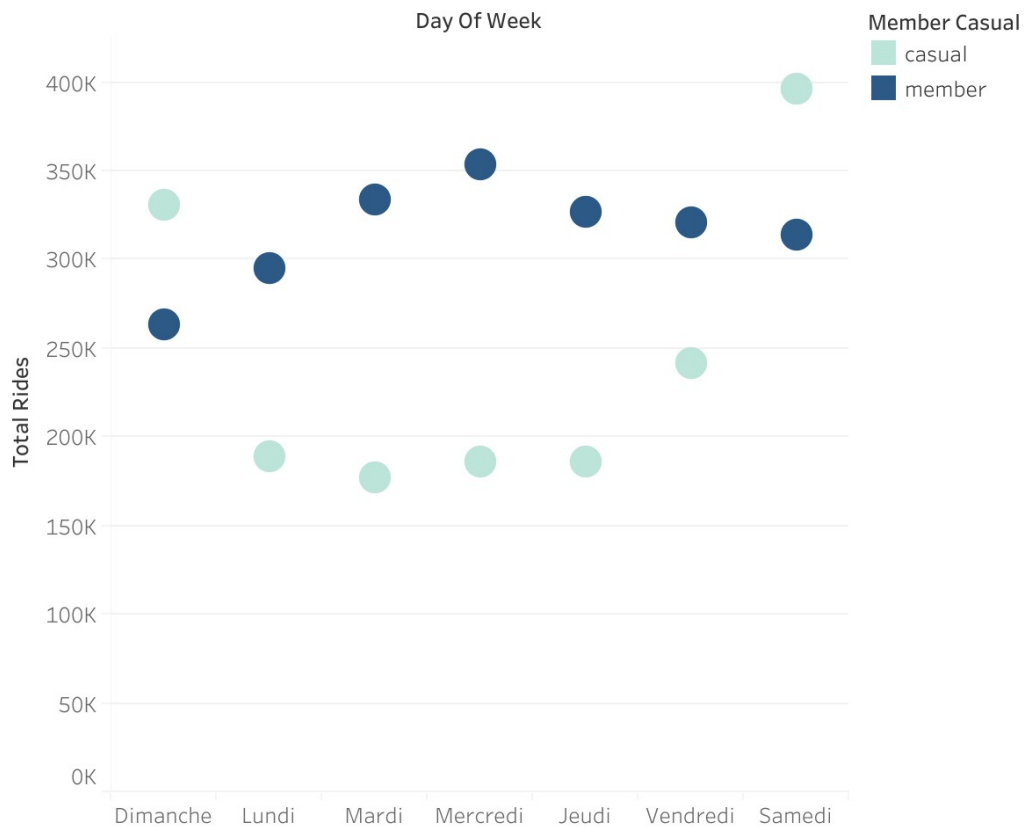
```
tripshour <- cleantrips1 %>% group_by(member_casual, starttime) %>%  
  summarise("Total Rides"=n())
```



**Calculate total rides by members and casuals on weekdays.** / Calculez le nombre total de trajets effectués par les membres annuels et les membres occasionnels en jour de la semaine.

```
tripsdays <- cleantrips1 %>%  
  group_by(member_casual, day_of_week) %>%  
  summarize("Total Rides" = n())
```

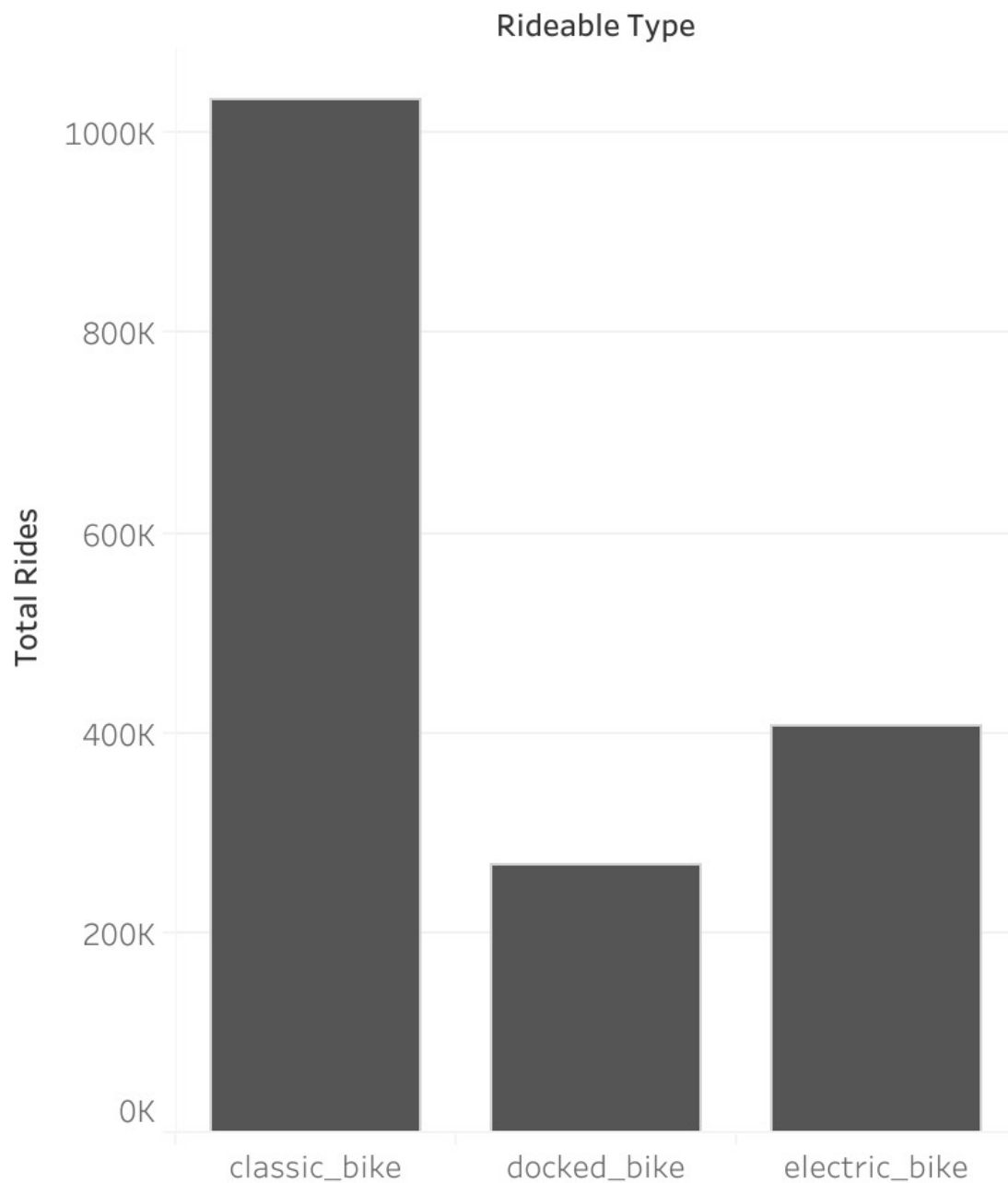
## Total Rides for Week days



**Calculation and visualisation of most used bike type by casual members.** / *Calcul et visualisation du type de vélo le plus utilisé par les membres occasionnels.*

```
ridetypecasual <- cleantrips1 %>%
  group_by(member_casual, rideable_type) %>% filter(member_casual
=="casual") %>%
  summarize("Total Rides" = n())
```

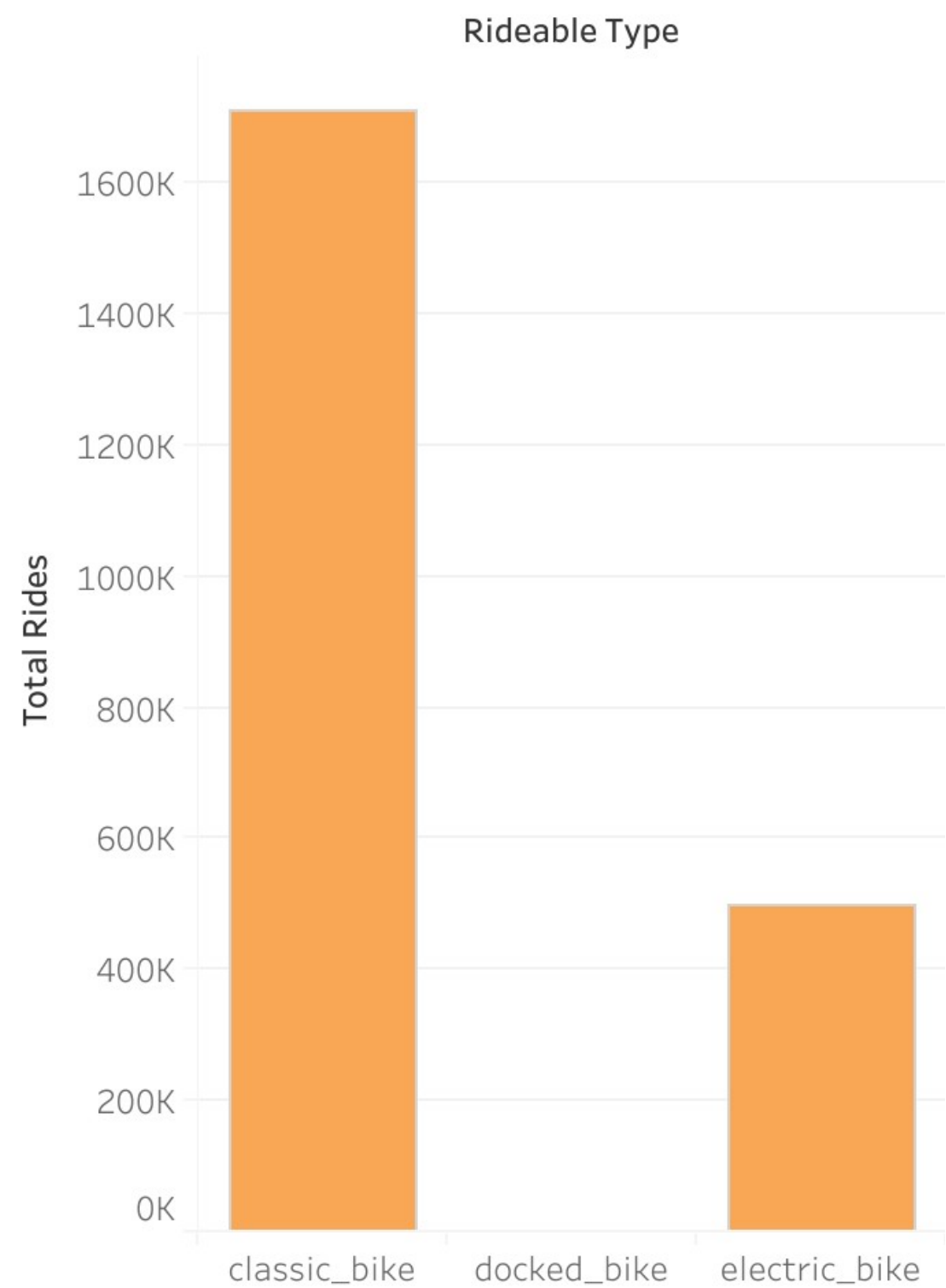
## Most Used Bike Type For Casuals



**Calculation and visualisation of most used bike type for annual members.** / *Calcul et visualisation du type de vélo le plus utilisé par les membres annuels.*

```
ridetypemember <- cleantrips1 %>%  
  group_by(member_casual, rideable_type) %>% filter(member_casual  
=="member") %>%  
  summarize("Total Rides" = n())
```

# Most Used Bike Type for Members

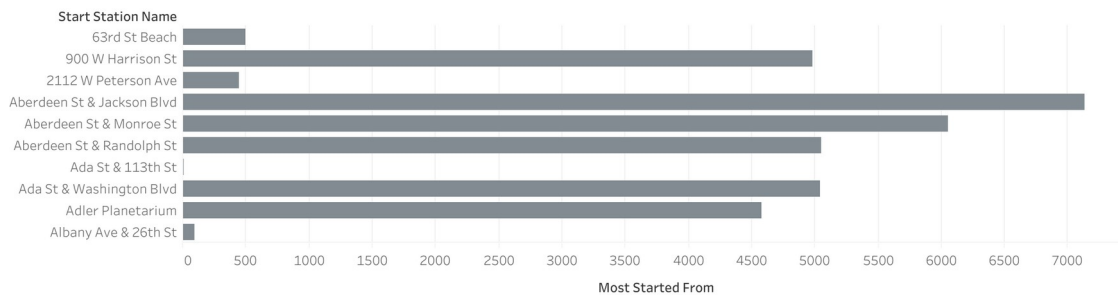




## Calculation and visualisation of the most used ten start stations -annual members / \*calcul et visualisation des dix stations de départ les plus utilisées -membres annuels\*\*

```
top10startmember <- cleantrips1 %>%  
  select(member_casual, start_station_name, day_of_week, starttime) %>%  
% filter(member_casual == "member") %>%  
  group_by(start_station_name) %>%  
  summarize("Most Started From" = n()) %>%  
  arrange(desc("Most Started From")) %>%  
  head(10)
```

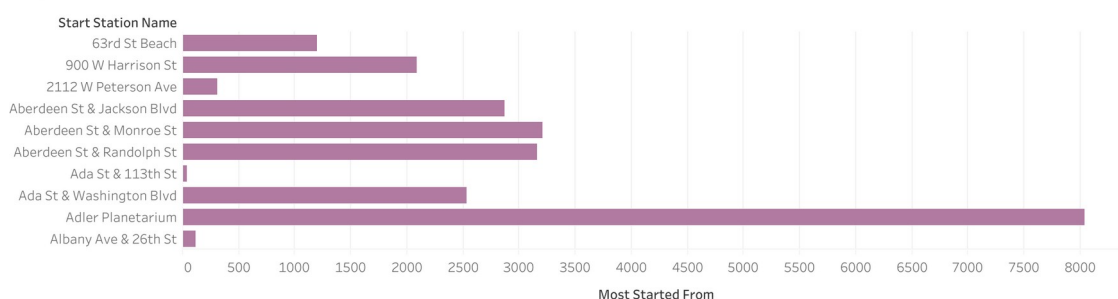
Top 10 Start Stations for Members



## Calculation and visualisation of the most used ten start stations -casual members / \*calcul et visualisation des dix stations de départ les plus utilisées -membres occasionnels\*\*

```
top10startcasual <- cleantrips1 %>%  
  select(member_casual, start_station_name, day_of_week, starttime) %>%  
% filter(member_casual == "casual") %>%  
  group_by(start_station_name) %>%  
  summarize("Most Started From" = n()) %>%  
  arrange(desc("Most Started From")) %>%  
  head(10)
```

Top 10 Start Stations for Casuals



## Share

**Now that i have performed my analysis and gained some insights into my data, i create visualizations to share my findings. / Maintenant que j'ai effectué mon analyse et obtenu un aperçu de mes données, je crée des visualisations pour partager mes découvertes.**

## Guiding questions / Questions d'orientation

- **Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?** / *Avez-vous pu répondre à la question de savoir comment les membres annuels et les cyclistes occasionnels utilisent différemment les vélos Cyclistic ?* Yes, I was able to find the answer to this question. For me, the main difference between annual members and casual members is the difference in their intended use. *Oui, j'ai pu trouver la réponse à cette question. Pour moi, la principale différence entre les membres annuels et les membres occasionnels est la différence dans leur utilisation prévue.*
- **What story does your data tell?** / *Quelle histoire racontent vos données ?* The story I have based on my data is how two separate members chose us independently of each other, their lifestyles and the region they are in. The key to this story is that they just have different purposes. *L'histoire que j'ai basée sur mes données est de savoir comment deux membres distincts nous ont choisis indépendamment l'un de l'autre, leurs modes de vie et la région dans laquelle ils se trouvent. Le point clé de cette histoire est qu'ils ont juste des objectifs différents.*
- **How do your findings relate to your original question?** / *Quel est le lien entre vos conclusions et votre question initiale ?* Based on my analysis, the connection between my first question and my conclusion is that we understand two different members better, so we can transform casual members to annual members by developing a strategy. *D'après mon analyse, le lien entre ma première question et ma conclusion est qu'on comprend mieux deux membres différents, donc on peut transformer des membres occasionnels en membres annuels en développant une stratégie.*
- **Who is your audience? What is the best way to communicate with them?** / *Quel est votre public ? Quelle est la meilleure façon de communiquer avec eux ?* At this stage my audiences are my manager Lily Moreno and Analytics Team. The best way to communicate with them is to create visualizations to share my findings. *À ce stade, mon public est ma responsable Lily Moreno et team d' Analytics. La meilleure façon de communiquer avec eux est de créer des visualisations pour partager mes découvertes.*

## ACT

**Now that i have finished creating my visualizations, i will act on my findings. I prepare the deliverables Morena asked me to create, including the three top recommendations based on my analysis. Guiding questions / Questions d'orientation**

- **What is your final conclusion based on your analysis?** / *Quelle est votre conclusion finale basée sur votre analyse ?* According to my analysis, main difference between casual members and annual members is just their habits. *Selon mon analyse, la principale différence entre les membres occasionnels et les membres annuels réside simplement dans leurs habitudes.*
- **How could your team and business apply your insights?** / *Comment votre équipe et votre entreprise pourraient-elles appliquer vos connaissances ?* By knowing the difference between members and understanding them, the marketing team could develop a marketing strategy to casual members. *En connaissant la différence entre*

*les membres et en les comprenant, l'équipe marketing pourrait développer une stratégie marketing auprès des membres occasionnels.*

- **What next steps would you or your stakeholders take based on your findings?** / *Quelles prochaines étapes vous ou vos managers entreprendriez-vous en fonction de vos conclusions ?* Based on my findings, we should create new opportunities for casual members, which can be helpful also for annual members. *Sur la base de mes conclusions, nous devrions créer de nouvelles opportunités pour les membres occasionnels, ce qui peut également être utile pour les membres annuels.*

## CONCLUSION

**According to my analysis** / *D'après mon analyse*

1. Casual members' average duration is more than annual members./ *La durée moyenne des membres occasionnels est supérieure à celle des membres annuels.*
2. Total rides per hour of each member type are close to each other./ *Le nombre total de trajets par heure de chaque type de membre est proche l'un de l'autre.*
3. Casual members and annual members mostly prefer classic bike. / *Les membres occasionnels et les membres annuels préfèrent majoritairement le vélo classique.*
4. The station where annual members' most started to take bikes is Aberdeen St & Jackson Boulevard/ *La station où les membres annuels ont le plus commencé à prendre des vélos est Aberdeen St & Jackson Boulevard*
5. The station where casual members' most started to take bikes is Adler Planetarium / *La station où les membres occasionnels ont le plus commencé à prendre des vélos est Adler Planetarium*
6. While casual members are prefer weekends, annual members prefer week days./ *Les membres occasionnels préfèrent les week-ends, alors que les membres annuels préfèrent les jours de semaine.*

-Since we do not have special data of users such as gender and age, it is not possible to separate two different member types in this sense. / *Étant donné que nous ne disposons pas de données spéciales sur les utilisateurs telles que le sexe et l'âge, il n'est pas possible de séparer deux types de membres différents dans ce sens.*

## MY SUGGESTIONS

- Since the docked bike is the least used type of bike, in this sense, if an extra budget is allocated, spending this budget by increasing the number of classic bikes or making broken classic bikes can benefit the company financially. / *Le vélo ancré étant le type de vélo le moins utilisé, en ce sens, si un budget supplémentaire est alloué, dépenser ce budget en augmentant le nombre de vélos classiques ou en fabriquant des vélos classiques cassés peut bénéficier financièrement à l'entreprise.*
- Since our goal is to convert occasional members to annual members, a more advantageous sales package/campaign can be prepared for weekends, but also covers weekdays./ *Puisque notre objectif est de convertir les membres occasionnels en membres annuels, un package/campagne de vente plus avantageux peut être préparé pour les week-ends, mais couvre également les jours de semaine.*

- By choosing the three stations where occasional members begin to use their bikes the most, advertising posters can be prepared on the new campaigns. These posters or ads can be use at bus stops and squares. */En choisissant les trois stations où les membres occasionnels commencent à utiliser le plus leur vélo, des affiches publicitaires peuvent être préparées sur les nouvelles campagnes. Ces affiches peuvent être accrochées aux arrêts de bus et sur les places.*

By Elifsu JAFFRES