2. Description des sources de données utilisées :

Il s'agit des données historiques de Cyclistic sur les déplacements à vélo des utilisateurs, contenant probablement les informations suivantes :

Le lien de téléchargement des données **:** Index of bucket "divvy-tripdata"

Elles ont organisé sur des formats csv par trimestre des données structurées. On prend les 4 trimestre de 2019 et le 1er trimestre de 2020 pour répondre à nos questions principales.

Selon l'outil utilisé on prend certains fichiers car c'est la partie gratuite.

Les variables des fichiers

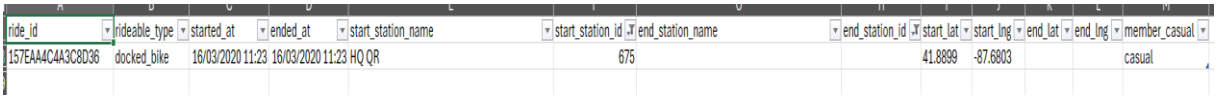
Variables du 1er , 2e, 3e et 4e trimestre 2019	**Variables de 1er trimestre 2020**
Trip_id	Ride_id
Start_time	Rideable_type
End_time	Started_at
bikeid	Ended_at
tripduration	Start_station_name
From_station_id	Start_station_id
From_station_name	End_station_id
to_station-id	End_station_name
To_station_name	Start_ing
usertype	Start_lat
gender	End_ing
birthyear	End_lat
	Member_casual

Le 1er trimestre 2019 contient des données sur les customer et les subscriber en total 365069. il y a 341906 subscriber et 23163 customer soit 93,66% subscriber et 6,34% customer.

Le 1er trimestre 2020 contient 426887 lignes dont 48480 clients occasionnels et 378407 membres réparties comme suit 11,31% et 88,68%.

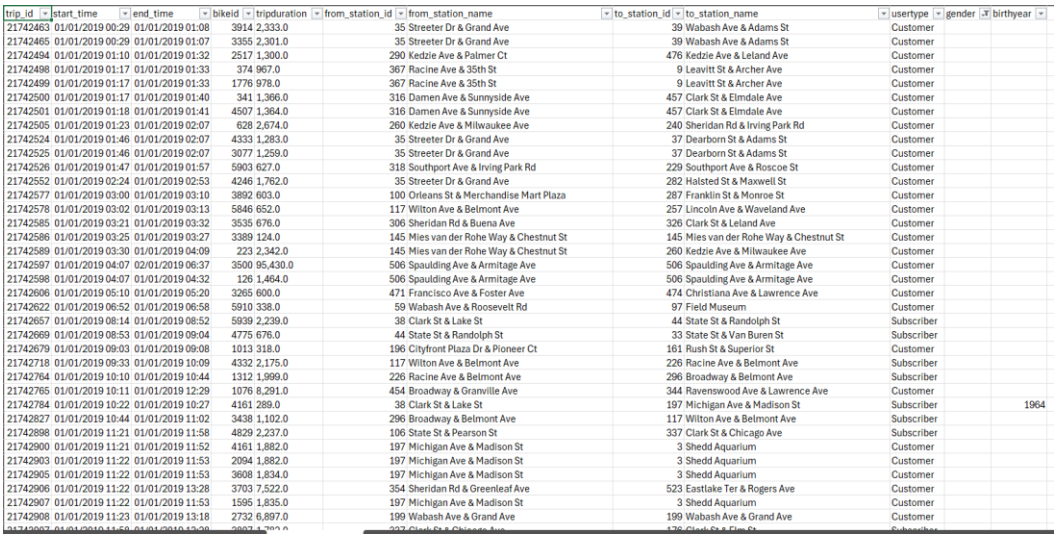
i. Vérifier s'il y a des données manquantes

- Dans le fichier trajet_2020_Q1 il y a quatre cellule vides dans les colonnes respectives end_station_name, end_station_id, end_lat, end_ing.



ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
157EAA4C4A3C8D36	docked_bike	16/03/2020 11:23	16/03/2020 11:23	HQ QR	675			41.8899	-87.6803			casual

- Le fichier trajet_2019_Q1, les colonnes genres et birthYear ont des cellules vides. La colonne genre contient 19711 cellules vides, et 18023 cellules vides pour la colonne birthyear.



trip_id	start_time	end_time	bikeid	tripduration	from_station_id	from_station_name	to_station_id	to_station_name	usertype	gender	birthyear
21742463	01/01/2019 00:29	01/01/2019 01:08	3914	2,333.0	35	Streeter Dr & Grand Ave	39	Wabash Ave & Adams St	Customer		
21742465	01/01/2019 00:29	01/01/2019 01:07	3355	2,301.0	35	Streeter Dr & Grand Ave	39	Wabash Ave & Adams St	Customer		
21742494	01/01/2019 01:10	01/01/2019 01:32	2517	1,300.0	290	Kedzie Ave & Palmer Ct	476	Kedzie Ave & Leland Ave	Customer		
21742498	01/01/2019 01:17	01/01/2019 01:33	374	967.0	367	Racine Ave & 35th St	9	Leavitt St & Archer Ave	Customer		
21742499	01/01/2019 01:17	01/01/2019 01:33	1776	978.0	367	Racine Ave & 35th St	9	Leavitt St & Archer Ave	Customer		
21742500	01/01/2019 01:17	01/01/2019 01:40	341	1,366.0	316	Damen Ave & Sunnyside Ave	457	Clark St & Elmdale Ave	Customer		
21742501	01/01/2019 01:18	01/01/2019 01:41	4507	1,364.0	316	Damen Ave & Sunnyside Ave	457	Clark St & Elmdale Ave	Customer		
21742505	01/01/2019 01:23	01/01/2019 02:07	628	2,674.0	260	Kedzie Ave & Milwaukee Ave	240	Sheridan Rd & Irving Park Rd	Customer		
21742524	01/01/2019 01:46	01/01/2019 02:07	4333	1,283.0	35	Streeter Dr & Grand Ave	37	Dearborn St & Adams St	Customer		
21742525	01/01/2019 01:46	01/01/2019 02:07	3077	1,259.0	35	Streeter Dr & Grand Ave	37	Dearborn St & Adams St	Customer		
21742526	01/01/2019 01:47	01/01/2019 01:57	5903	627.0	318	Southport Ave & Irving Park Rd	229	Southport Ave & Roscoe St	Customer		
21742552	01/01/2019 02:24	01/01/2019 02:53	4246	1,762.0	35	Streeter Dr & Grand Ave	282	Halsted St & Maxwell St	Customer		
21742577	01/01/2019 03:00	01/01/2019 03:10	3892	603.0	100	Orleans St & Merchandise Mart Plaza	287	Franklin St & Monroe St	Customer		
21742578	01/01/2019 03:02	01/01/2019 03:13	5846	652.0	117	Wilton Ave & Belmont Ave	257	Lincoln Ave & Waveland Ave	Customer		
21742585	01/01/2019 03:21	01/01/2019 03:32	3535	676.0	306	Sheridan Rd & Buena Ave	326	Clark St & Leland Ave	Customer		
21742586	01/01/2019 03:25	01/01/2019 03:27	3389	124.0	145	Mies van der Rohe Way & Chestnut St	145	Mies van der Rohe Way & Chestnut St	Customer		
21742589	01/01/2019 03:30	01/01/2019 04:09	223	2,342.0	145	Mies van der Rohe Way & Chestnut St	260	Kedzie Ave & Milwaukee Ave	Customer		
21742597	01/01/2019 04:07	02/01/2019 06:37	3500	95,430.0	506	Spaulding Ave & Armitage Ave	506	Spaulding Ave & Armitage Ave	Customer		
21742598	01/01/2019 04:07	01/01/2019 04:32	126	1,464.0	506	Spaulding Ave & Armitage Ave	506	Spaulding Ave & Armitage Ave	Customer		
21742606	01/01/2019 05:10	01/01/2019 05:20	3265	600.0	471	Francisco Ave & Foster Ave	474	Christiana Ave & Lawrence Ave	Customer		
21742622	01/01/2019 06:52	01/01/2019 06:58	5910	338.0	59	Wabash Ave & Roosevelt Rd	97	Field Museum	Customer		
21742657	01/01/2019 08:14	01/01/2019 08:52	5939	2,239.0	38	Clark St & Lake St	44	State St & Randolph St	Subscriber		
21742669	01/01/2019 08:53	01/01/2019 09:04	4775	676.0	44	State St & Randolph St	33	State St & Van Buren St	Subscriber		
21742679	01/01/2019 09:03	01/01/2019 09:08	1013	318.0	196	Cityfront Plaza Dr & Pioneer Ct	161	Rush St & Superior St	Customer		
21742718	01/01/2019 09:33	01/01/2019 10:09	4332	2,175.0	117	Wilton Ave & Belmont Ave	226	Racine Ave & Belmont Ave	Subscriber		
21742764	01/01/2019 10:10	01/01/2019 10:44	1312	1,999.0	226	Racine Ave & Belmont Ave	296	Broadway & Belmont Ave	Subscriber		
21742765	01/01/2019 10:11	01/01/2019 12:29	1076	8,291.0	454	Broadway & Granville Ave	344	Ravenswood Ave & Lawrence Ave	Customer		
21742784	01/01/2019 10:22	01/01/2019 10:27	4161	289.0	38	Clark St & Lake St	197	Michigan Ave & Madison St	Subscriber		1964
21742827	01/01/2019 10:44	01/01/2019 11:02	3438	1,102.0	296	Broadway & Belmont Ave	117	Wilton Ave & Belmont Ave	Subscriber		
21742898	01/01/2019 11:21	01/01/2019 11:58	4829	2,237.0	106	State St & Pearson St	337	Clark St & Chicago Ave	Subscriber		
21742900	01/01/2019 11:21	01/01/2019 11:52	4161	1,882.0	197	Michigan Ave & Madison St	3	Shedd Aquarium	Customer		
21742903	01/01/2019 11:22	01/01/2019 11:53	2094	1,882.0	197	Michigan Ave & Madison St	3	Shedd Aquarium	Customer		
21742905	01/01/2019 11:22	01/01/2019 11:53	3608	1,834.0	197	Michigan Ave & Madison St	3	Shedd Aquarium	Customer		
21742906	01/01/2019 11:22	01/01/2019 13:28	3703	7,522.0	354	Sheridan Rd & Greenleaf Ave	523	Eastlake Ter & Rogers Ave	Customer		
21742907	01/01/2019 11:22	01/01/2019 11:53	1595	1,835.0	197	Michigan Ave & Madison St	3	Shedd Aquarium	Customer		
21742908	01/01/2019 11:23	01/01/2019 13:18	2732	6,897.0	199	Wabash Ave & Grand Ave	199	Wabash Ave & Grand Ave	Customer		

Etudier l'impartialité de nos données en basant sur la méthode ROCCC et ses crédibilités en étudiant les différents types de biais.

Les données disponibles possèdent les variables nécessaires pour répondre aux questions principales sur l'utilisation des vélos, la durée du trajet, le type d'utilisateur, le nom et l'identifiant des stations de départ et d'arrivée par les clients occasionnels et les membres. Elles sont incomplètes, donc dans la partie de nettoyage, nous allons compléter les cellules qui peuvent se compléter ou supprimer.

L'aspect éthique des fichiers sont respectés car nous ne pouvons pas relier les informations personnelles identifiables des cyclistes

Questions directives	
Où se trouvent les données	Les données qu'on va utiliser se trouvent dans la base de données interne à l'entreprise Cyclitic avec ce lien : Index of bucket "divvy-tripdata"
Comment les données sont-elles organisées ?	Elles sont organisées par des fichiers zippés par mois de chaque année, les dates de modifications ou création de chaque fichier, son poids. Les fichiers sont en csv.
Ces données présentent-elles des problèmes de partialité ou de crédibilité ? Vos données sont-elles conformes à la ROCCC ?	- Les données sont impartiales, car elles contiennent les informations qui peuvent répondre à nos interrogations. - Elles sont intègres, car elles valident le processus ROCCC : données complètes et exactes, sources données interne à l'entreprise c-à-d fiable, données pertinentes à notre objectif, données ne sont pas erronées, et on a la source de données qui est l'archive de toutes les réservations des usagers.

Comment abordez-vous les questions de licence, de confidentialité, de sécurité et d'accessibilité ?	L'entreprise Cyclitic met en place des mécanismes pour respecter les données de ses clients à savoir la confidentialité, le consentement, le droit de supprimer ses données
Comment avez-vous vérifié l'intégrité des données ?	La vérification de l'intégrité de la base de données se fait par le processus ROCCC et l'étude de biais
Comment cela vous aide-t-il à répondre à votre question	La vérification de l'intégrité des données assure la fiabilité des données, et qu'on sache bien que notre analyse et conclusion seront justes et pertinentes

Il y a des données manquantes dans notre feuille de données, plus spécifiquement aux colonnes gender et birthyear. Le pourcentage de cellules vides dans la colonne genre est de 5,40% soit 19711, ce nombre n'est pas insignifiant. Pour ne pas avoir une base incomplète, et en plus ces variables ne sont pas pertinentes pour répondre à notre problématique.

Par conséquent j'ai supprimé ces 2 colonnes dans notre base de données.