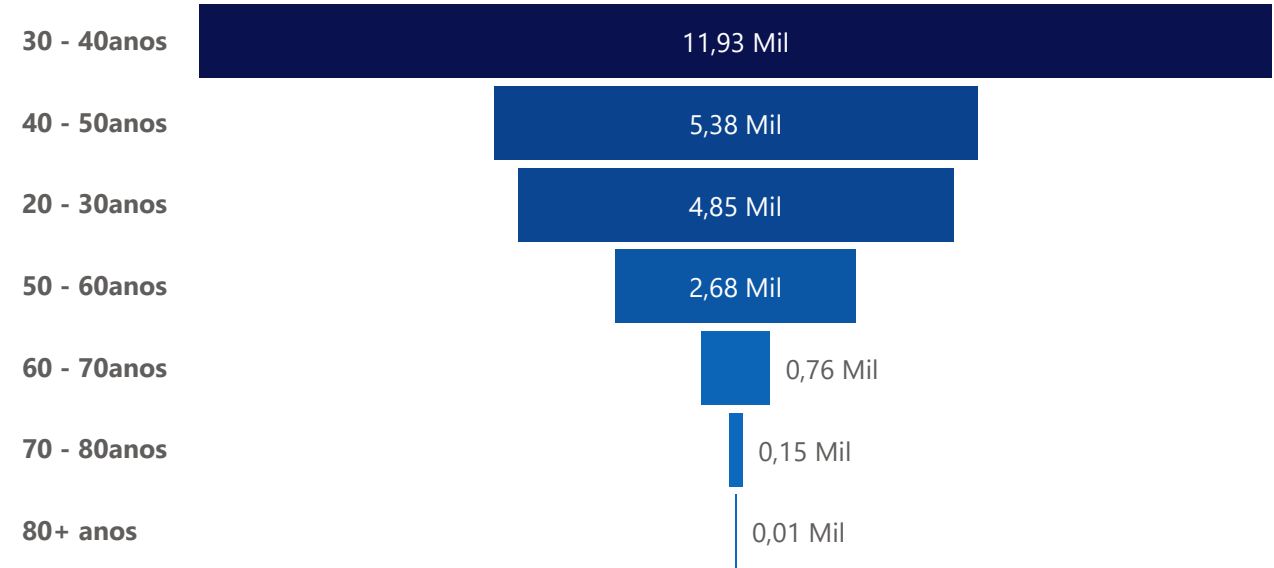
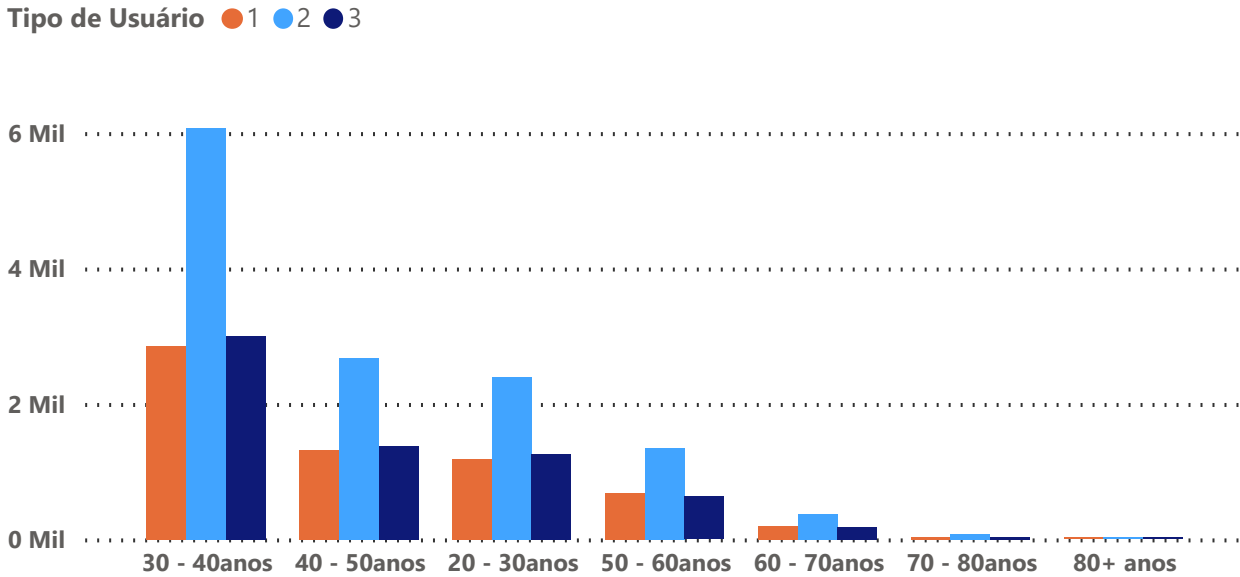


Contagem de corridas por faixa de idade



Contagem de corridas por tipo de usuário e idade



- 1 - Corridas Grátis
- 2 - Corridas pagas por viagem
- 3 - Corridas feitas por pagadores mensais

Estações de partida mais utilizadas (Contagem de corridas)

San Francisco Caltrain Station 2 (Townsend St at 4th St)	3635
Berry St at 4th St	3030
Montgomery St BART Station (Market St at 2nd St)	2998

Contagem de corridas por gênero



```
In [ ]: import pandas as pd

ride_sharing = pd.read_csv('ride_sharing_new.csv')
ride_sharing.head()
```

```
In [ ]: #Retirando a string minutes da coluna duration

ride_sharing['duration'] = ride_sharing['duration'].str.strip('minutes')

#transformando o tipo da coluna duration para int e verificando

ride_sharing['duration'] = ride_sharing['duration'].astype('int')
ride_sharing.head()
ride_sharing.info()
```

```
In [ ]: #Os dados de tipo de usuario sao categoricos, por isso transforma-los em category e verificar

ride_sharing['user_type'] = ride_sharing['user_type'].astype('category')
ride_sharing['user_type'].describe()
```

```
In [ ]: #analizando a coluna user_birth_year

ride_sharing['user_birth_year'].info()
ride_sharing['user_birth_year'].min()

#definido que o ano de nascimento de um usuario deve ser nao menor que 1930

ride_sharing.loc[ride_sharing['user_birth_year'] < 1930 , 'user_birth_year'] = 1930

#verificando ausencia de valores

ride_sharing.isna().sum()
```

```
In [ ]: #criando grupos da coluna duration

import numpy as np
```