Recreational Park Management

**Group 6 / Topic 4:**

Jaynil Jaiswal (17114040)

Peddi Krishna Yeswanth (17114055)

Balwant Singh (17114018)

Akhilesh Kumar (17114007)

Ajay Dayma (17114006)

# Problem Statement:

Data mining on Recreational Park Management. This includes different management related details like Expenditure, Revenue, Events, Ratings ... etc.

# Datasets:

We are using three different types of data sets in this project, each of which is related to one of the above mentioned details. The amount of open source data available is very sparse so we used the amount of data available to extract meaning rules/ knowledge.

* **Event Types Dataset (Dataset 1):-** This data is collected from the official website of New York City Department of Parks and Recreation (NYC park management). It contains data related to different events that are being held in the span of one year (2018-19).

It contains a total of 12 fields related to the event.

* Unit: This describes about the name of unit that performed/organized the event
* Group Name/Partner: This describes about the name of group that requested the event or partner that helped organize the event
* Date and Time: Date and time of the event
* Borough: This describes about the Borough in which the event was held
* LocationType: This describes the type of the park where the event is held.
* Location: Exact Location in the Borough where the event is held
* Event Name: Name of the event
* Event Type: Type of the event that is being held
* Category: Type of activity performed at the event
* Classification: Further describes about the event based on category
* Audience: Type of population participating in the event
* Attendance: Number of participants who attended the event
* **Parks Budget Dataset (Dataset 2):-** This data is collected from the official data portal of Las Vegas. It was data of a combined budget on every department of the city. Our topic aligns with the department of “Parks and Recreation” data. It contains data related to different types of expenditures/revenues and their amount.

It contains a total of 12 fields in the dataset.

* id: Primary key of data
* Fiscal\_year
* Function: Type of functions happening at parks. Mostly culture & recreation.
* Program: NULL
* Department: Parks and Recreation
* Type: General Category of expenditure
* Type\_Detail: Detailed Category of expenditure
* Approved\_amount: Amount of money
* Fund\_type: General or Special Revenue
* exFund: General or Special Revenue
* ExpenseType: Expenditure or Revenue
* ObjectId: Foreign key to Original Dataset
* **Parks User Satisfaction Survey for Different Event Types Dataset (Dataset 3):-** This data is also collected from the official data portal of Las Vegas.

It contains a total of 11 fields out of which 5 fields are useful.

* ResponseID
* EndDate
* Satisfaction\_Level : 'Poor':0, 'Very Dissatisfied':1, 'Dissatisfied':2, 'Neutral':3, 'Fair':4, 'Satisfied':5, 'Very Satisfied':6, 'Good':7
* Program: Event Type
* FY: Fiscal Year

# Data pre processing:

## Dataset 1:-

Firstly, we analyzed the data to remove redundant columns like Unit, Group Name, Date and Time. Then we used the combination of (Event Name, Event Type, Category) to categorize the classes and the events in different classes related to these. After pre-processing, the data contains Borough, Class of Event(extracted from above combination of event information), Audience and Attendance.

* Borough Types: Bronx, Brooklyn, Manhattan, Queens, Staten Island
* Class Types: AC&F(Arts & Culture & Fun), Awards, Dance, Fitness, Lecture/Panel/Talk, M.U.T.S(Movie Under the Sky), Movies, Music, Play, Sport, Others.
* Audience: Tot(Toddler), Children, Teens, Young Adult, Adults, Seniors
* Attendance: No of attendees.

## Dataset 2:-

The only preprocessing required was:

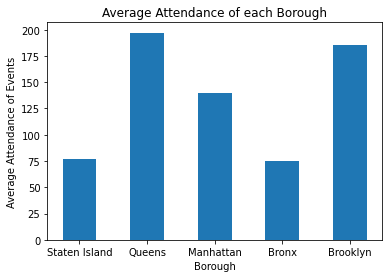
* To get department “parks and recreation”
* Convert dataframe to multiindex dataframe for plotting

## Dataset 3:-

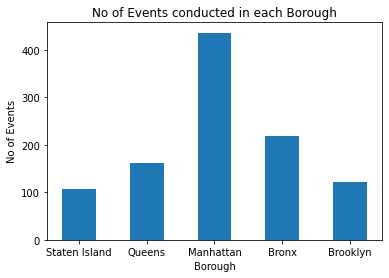
Converting Satisfaction Level to ordered data(0-7 stars).

# Results:

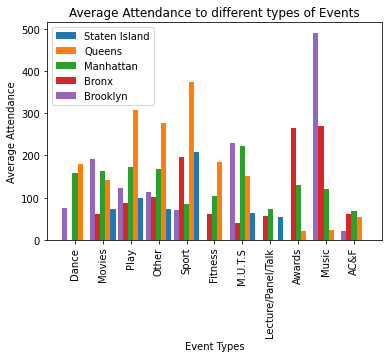
## Dataset 1:



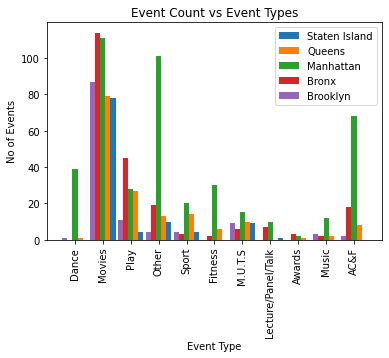
As we can clearly see, the average amount of Attendance at Queens, Brooklyn regions is much higher compared to other regions and is least in Staten Island. The above result is as expected as the population of Queens, Brooklyn is twice compared to other cities and is least in Staten Island[[1]](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City#Background).



As we can clearly see, Manhattan has the highest no of events being held in the given year 2018-19. This can be inferred from the fact that Manhattan is the borough with the highest population density, more than twice when compared to other boroughs[[2]](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City#Description_of_the_boroughs).



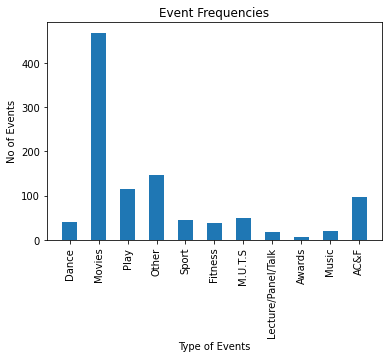
As we can clearly see, Brooklyn has the highest Average Attendance for Music events. This can be evident from the fact that Brooklyn, which is known for its cultural, social, and ethnic diversity. We can observe that Manhattan, Staten Island have a smooth distribution of Attendance over all types of events.



As we can clearly see, the events with movie type drastically outclasses the other type of events, the no of events held in Manhattan Outclasses other. This can be evident from that fact, Manhattan has more than twice the population when compared to other boroughs. The US is one of the countries which streams most videos online/ watch movies[[3]](https://www.statista.com/statistics/272835/share-of-internet-users-who-watch-online-videos/).

**Most Frequent Itemsets:**

We are calculating frequent itemsets with candidate items as the type of events and the type of Audience participating in these events to get the idea of which type of event attracts what age group of people, so that we estimate Audience during the next event. These itemsets can be used to generate association rules, which gives us a better understanding of what type of event attracts which age group, which age group of people appear in pairs/ groups.



As we can see, the frequency of most of the event types are less than 100, most prevailing events are Movies, Others, Play. We are considering a minimum support count of 100, to calculate the frequent item sets. We are generating rules with a minimum confidence level of 0.3 (so as to allow event types to occur, as the no of Audience far exceed event types).

Rules: ('Play',): (('Children',), 0.9217391304347826), ('Children', 'Other'): (('Teens',), 0.9),

('Other', 'Tot'): (('Children',), 0.9807692307692307),

('Adults', 'Other'): (('Children', 'Teens'), 0.8695652173913043),

('Adults', 'Other', 'Seniors'): (('Teens',), 0.9532710280373832),

('Adults', 'Other', 'Young Adult'): (('Teens',), 0.9716981132075472),

('Other', 'Seniors', 'Young Adult'): (('Adults',), 0.9901960784313726),

('Other', 'Teens'): (('Children',), 0.8925619834710744),

('Adults', 'Other', 'Teens'): (('Children',), 0.9433962264150944),

('Other', 'Seniors', 'Teens'): (('Adults',), 1.0),

('Children', 'Other', 'Teens'): (('Adults',), 0.9259259259259259),

('Children', 'Other', 'Young Adult'): (('Adults', 'Teens'),

0.9523809523809523)

We can see other trivial rules like ('Adults', 'Movies', 'Tot', 'Young Adult'): (('Children', 'Seniors', 'Teens'),1.0),

('Movies', 'Seniors', 'Tot', 'Young Adult'): (('Adults', 'Children', 'Teens'),1.0),

('Adults', 'Movies', 'Teens', 'Tot'): (('Seniors',), 1.0),

('Adults', 'Movies', 'Teens', 'Young Adult'): (('Children', 'Seniors', 'Tot'),1.0),

('Movies', 'Teens', 'Tot', 'Young Adult'): (('Adults', 'Children', 'Seniors'),1.0),

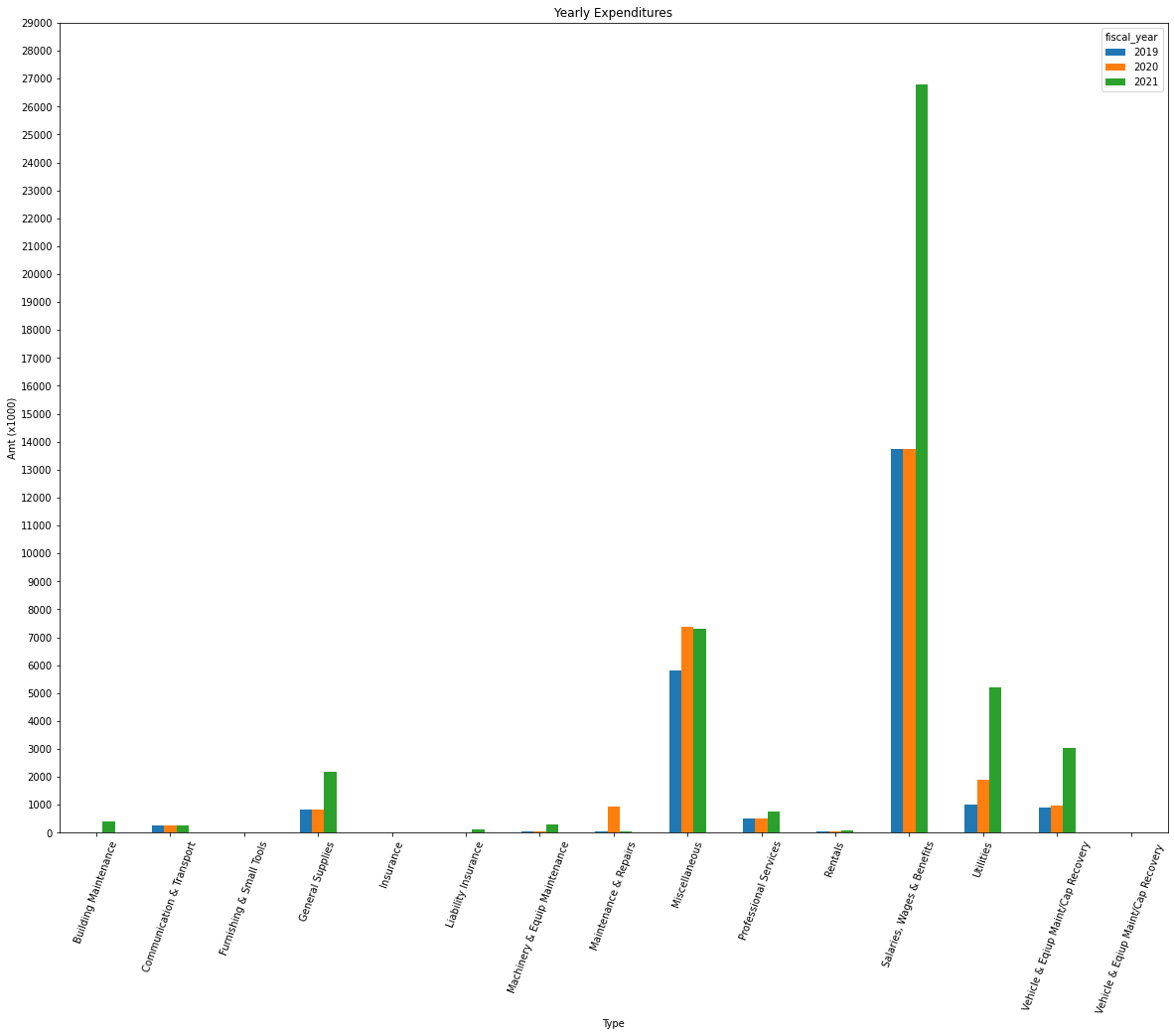
('Movies', 'Seniors', 'Teens', 'Young Adult'): (('Adults', 'Children', 'Tot'),1.0),

('Movies', 'Seniors', 'Teens', 'Tot'): (('Adults',), 1.0),

('Adults', 'Movies', 'Seniors', 'Teens', 'Tot'): (('Children', 'Young Adult'),1.0)

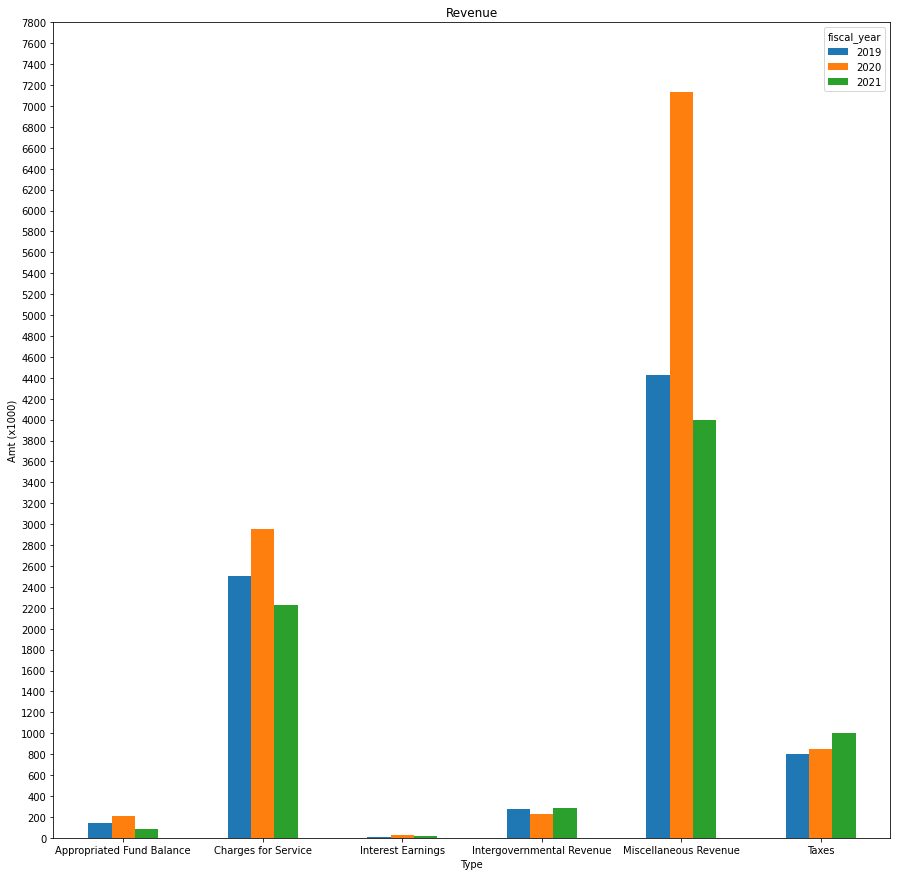
Results of Budget Dataset:

Annual Expenditure Review:



As we clearly see, (Salaries, Wages & Wages) contribute to most of the Annual amount (scaled by 1000) and we can see a drastic increase in 2021 due to safety measures issued due to COVID.

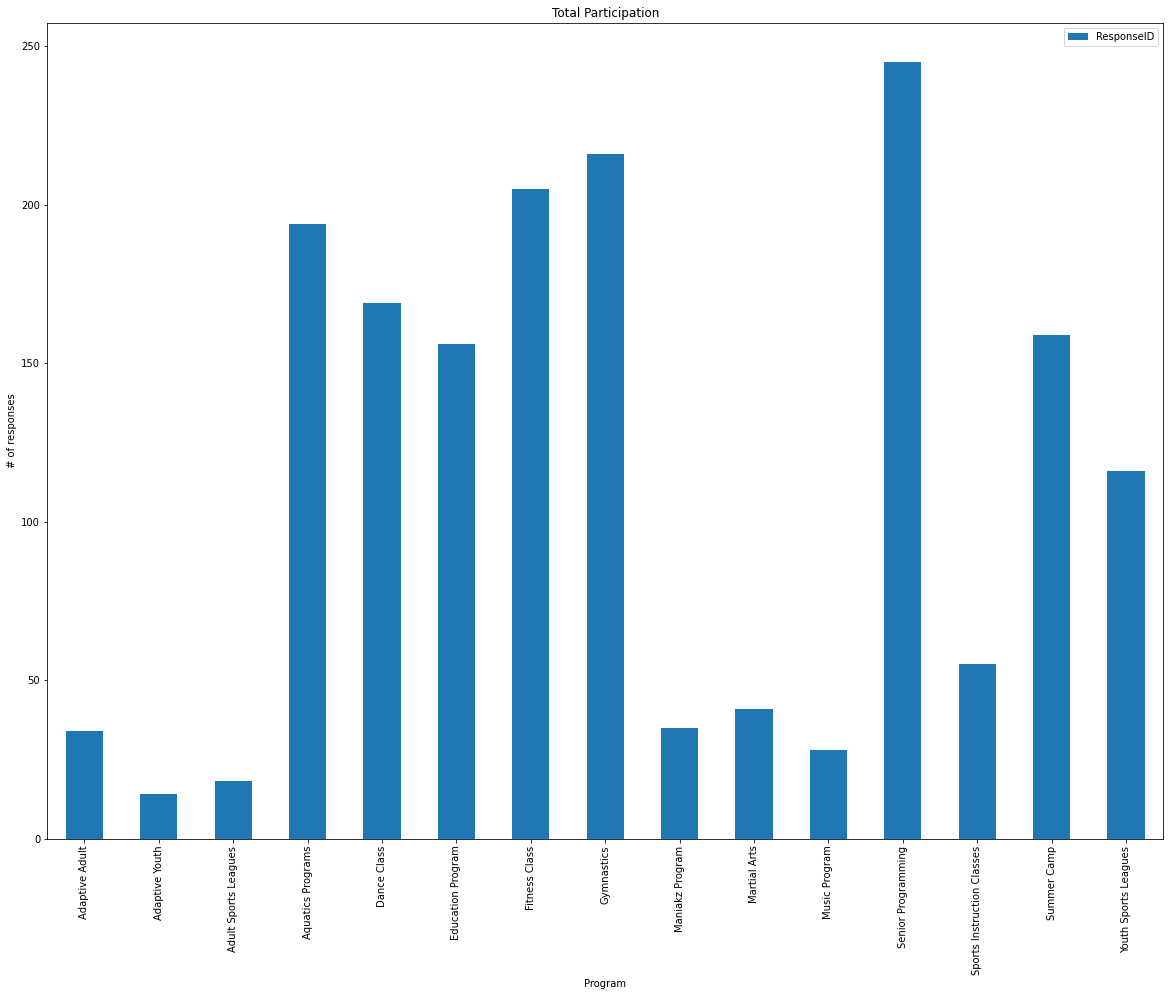
Annual Revenue Review:



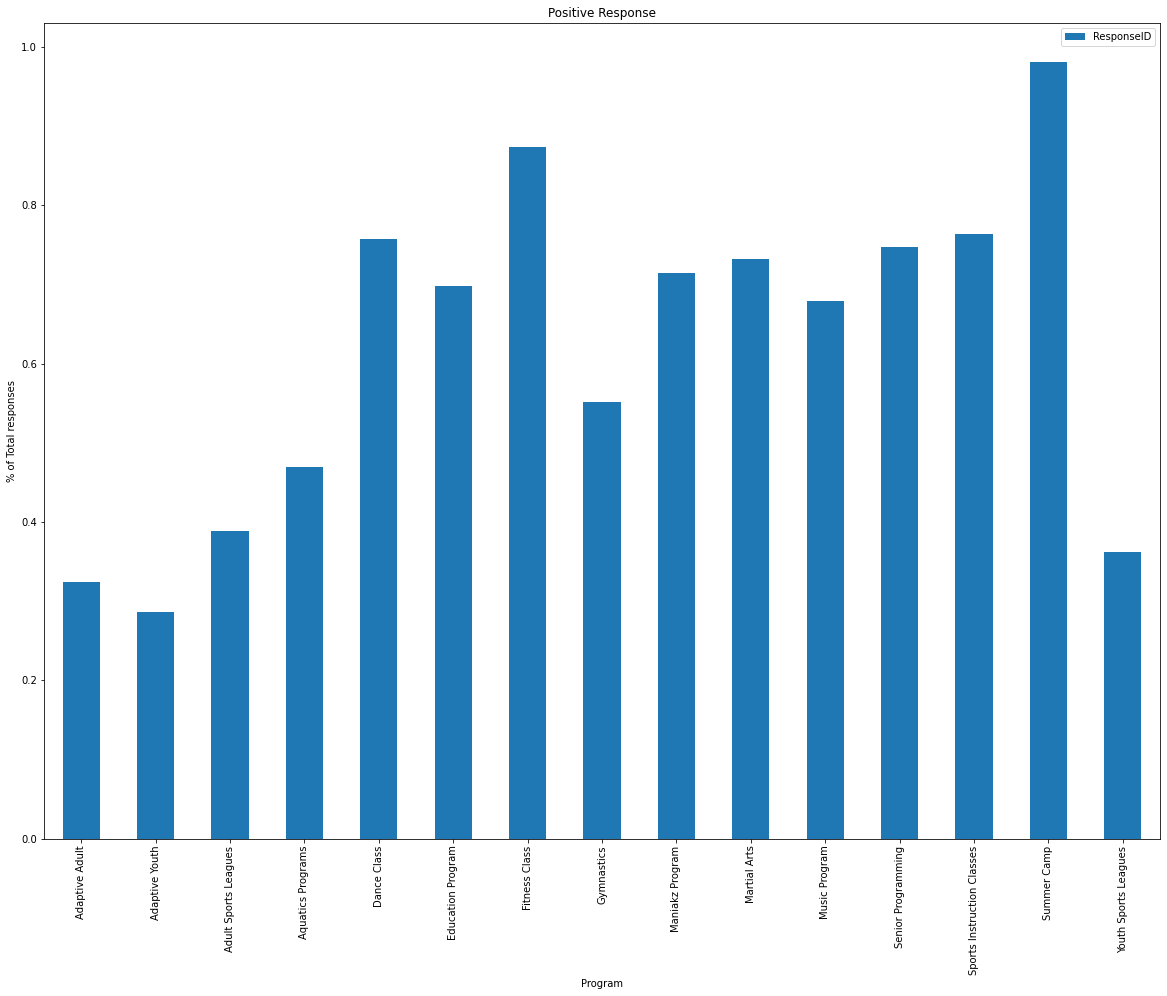
As we can see, Miscellaneous Revenue (which mainly includes rentals, donations and sponsorships) increased a lot in 2020. During COVID, there is a lot of money flowing through Donations, SponsorShips.

Results of User Satisfaction of Event types:

Total Participation:

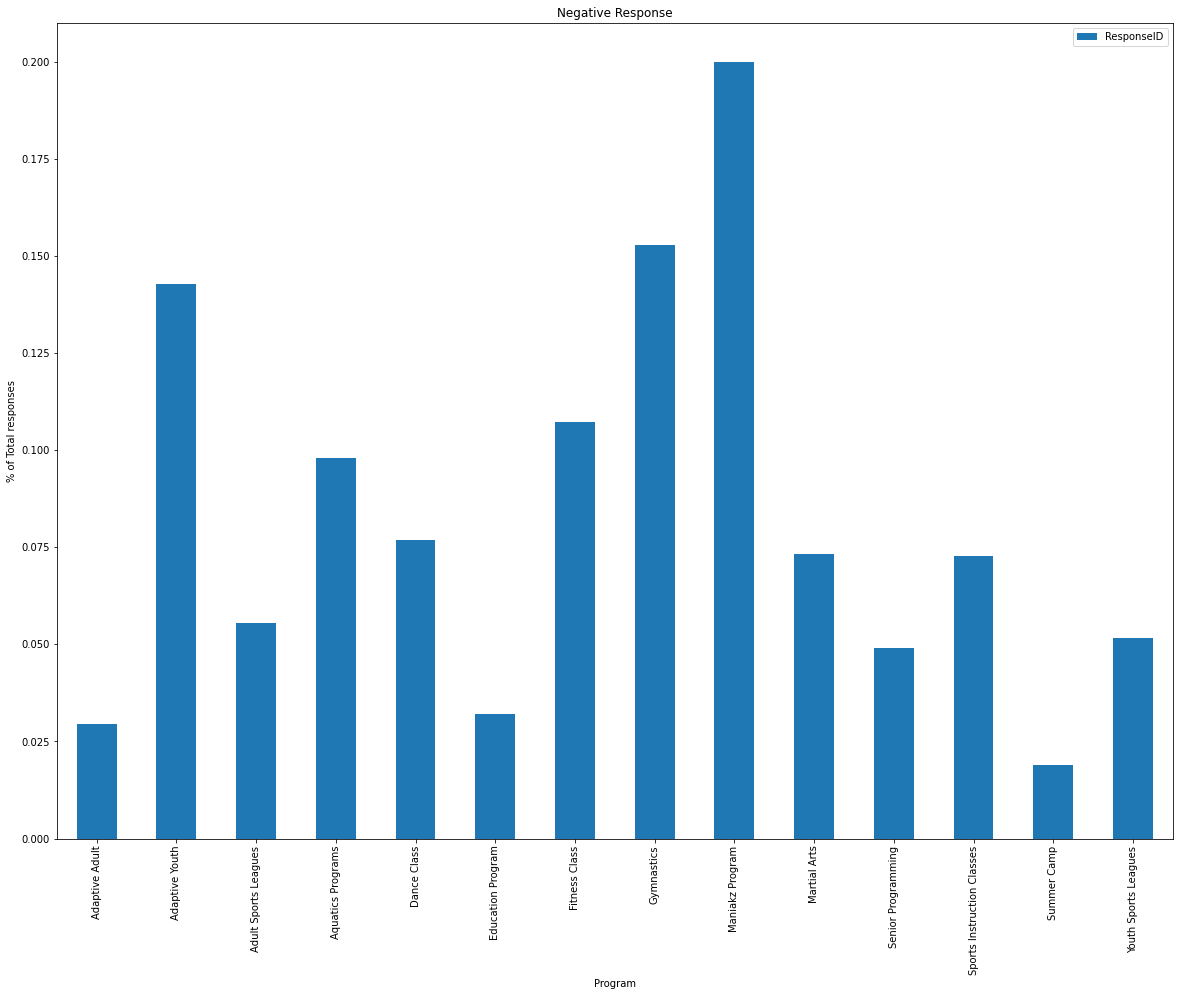


Positive Review:



This plot gives us the event types which were positively received by the audience which participated in the event. As we can see fitness and summer camp are more popular.

Negative Review:



Similarly we see that the Maniakz program was not popular among the audience.