AICP Internship Task

App User Segmentation

In the problem of app user segmentation, we need to group users based on how they engage with the app. We have data about the users based on how they engage with the app.

Find dataset "**userbehaviour.csv**"

Below are all the features in the dataset:

1. **userid:** The identity number of the user;
2. **Average Screen Time:** The average screen time of the user on the application;
3. **Average Spent on App (INR):** The average amount spent by the user on the application;
4. **Left Review:** Did the user leave any reviews about the experience on the application? (1 if true, otherwise 0)
5. **Ratings:** Ratings given by the user to the application;
6. **New Password Request:** The number of times the user requested a new password;
7. **Last Visited Minutes:** Minuted passed by when the user was last active;
8. **Status:** Installed if the application is installed and uninstalled if the user has deleted the application;

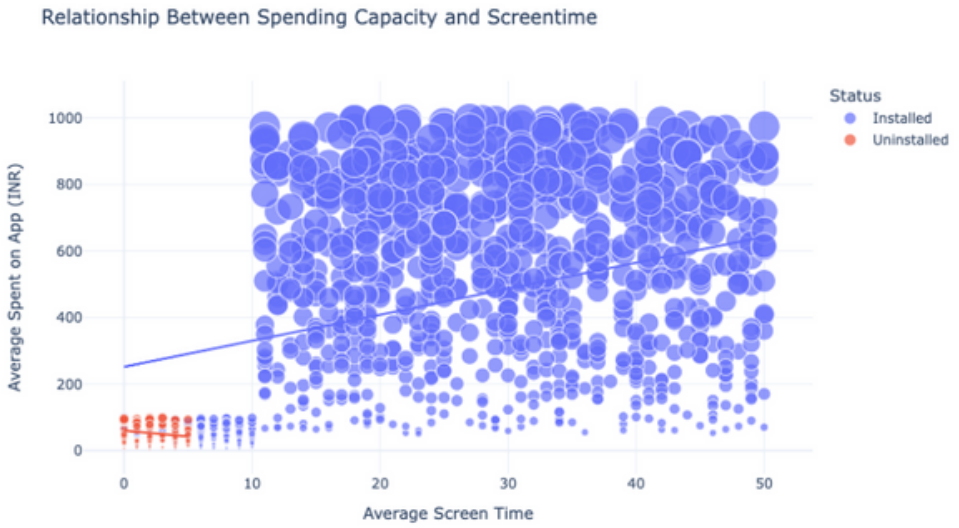
Find relationships between the users who are still using the application and the users who have uninstalled the application and create user segments to understand the retained users and the users that can be retained before they move to other alternatives.

Q.1: Import data and check null values, column info, and descriptive statistics of the data.

Q.2: Check the highest, lowest, and average screen time of all the users.

Q.3: Check the highest, lowest, and the average amount spent by all the users.

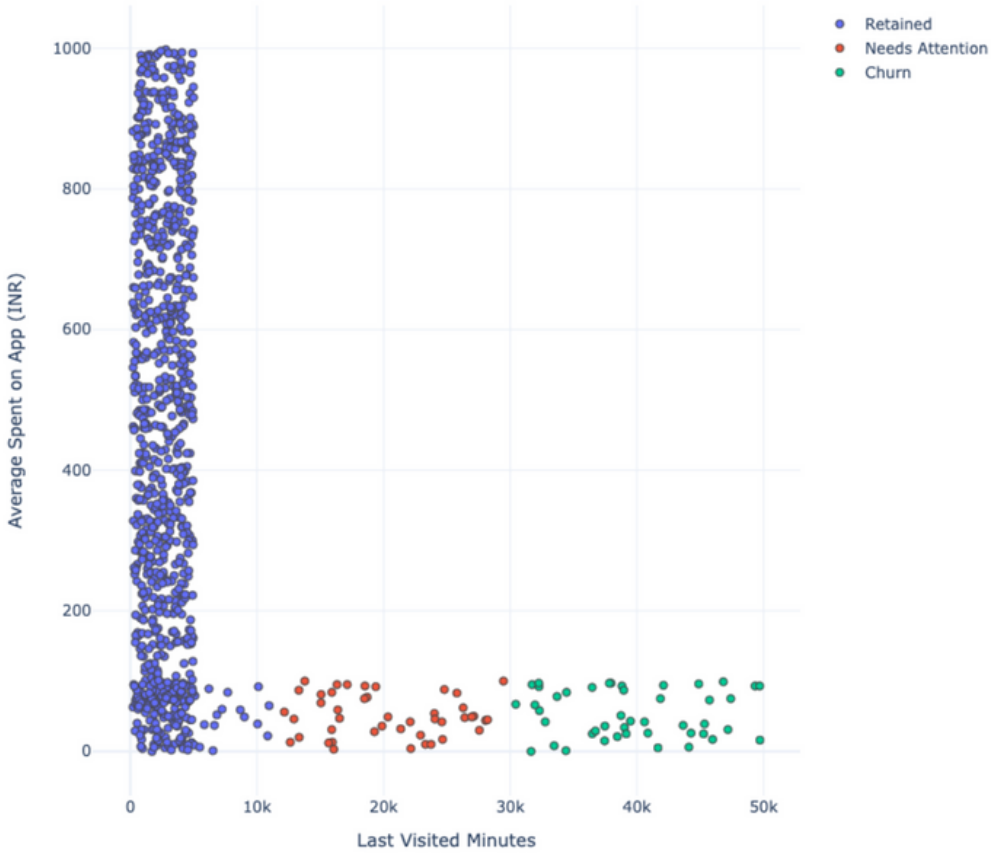
Q.4: Now check the relationship between the spending capacity and screen time of the active users and the users who have uninstalled the app. Also explain your observation.



Q.5: Now check the relationship between the ratings given by users and the average screen time. Also explain your observation.

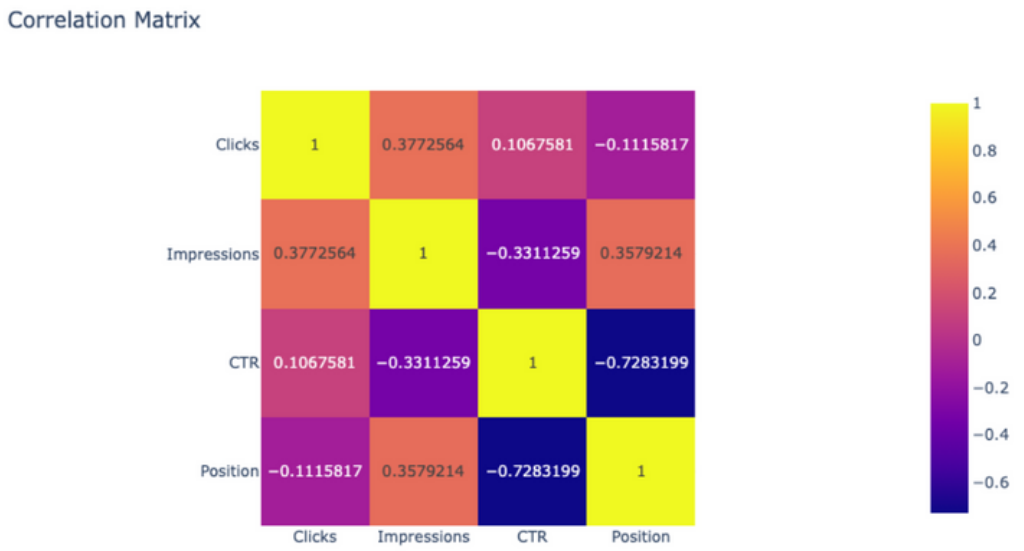
Q.6: Now move forward to App User segmentation to find the users that the app retained and lost forever. You can use the K-means clustering algorithm in Machine Learning for this task. Also, tell the number of segments you have got.

Q.7: Now visualize the segments.



Q.8: Explain the summary of your working.

Q.6: Now check the correlation between different metrics. Also explain your observation from the correlation matrix



Q.7: Now, detect anomalies in search queries. You can use various techniques for anomaly detection. A simple and effective method is the Isolation Forest algorithm, which works well with different data distributions and is efficient with large datasets.

Show results like this

	Top queries	Clicks	Impressions	CTR	Position
0	number guessing game python	5223	14578	0.3583	1.61
1	thecleverprogrammer	2809	3456	0.8128	1.02
2	python projects with source code	2077	73380	0.0283	5.94
4	the clever programmer	1931	2528	0.7638	1.09
15	rock paper scissors python	1111	35824	0.0310	7.19
21	classification report	933	39896	0.0234	7.53
34	machine learning roadmap	708	42715	0.0166	8.97
82	r2 score	367	56322	0.0065	9.33
167	text to handwriting	222	11283	0.0197	28.52
929	python turtle	52	18228	0.0029	18.75