Good afternoon everyone, we are team Stats trek. Today we are going to present our findings on the Expedia data. The main focus of our research is to provide customized hotel recommendations based on past booking history. We hope that a user-friendly service can increase the booking rate of Expedia users in United States.

First of all, we engineered new features to enhance our machine-learning algorithm. The three most significant variables are days prior to travel, travel time, and international. We believe these three features will **reveal** some **behavior differences** **among** the Expedia users.

Then, we removed some abnormal data, such as those bookings with later check-in date than check out date, which doesn’t make sense right? After cleaning, we extracted variables of interest and applied the clustering algorithm k-means to divide users into three groups. We find three variables are most significant, days prior to travel, number of adults and travel time. By applying descriptive statistics, we **identified** the **characteristics** of three groups of travellers pertaining to booking behavior.

This is the table from cluster analysis. All the numbers are standardized for better comparability.

The first category is international vacationers. They are the early birds when booking hotels and they have the longest travel time.

The second group is business professionals, who usually travel alone and book with mobile devices. Unlike the previous vacationers, they have the least time to prepare the trip and have the shortest length of stay.

The last one is family travelers. They travel in large groups with children and tend to book relatively ahead.

Next my teammate Jiahao will provide deeper insight into the data.

Recommendations:

Based on k-means analysis, we generate this 3d scatter plot to better visualize our clustering result. The red one represents international vacationers, they are the largest group among three. The blue points are the business people. They are at the corner of the graph, and they constitute the smallest proportion of all the travelers. The green one is the family group. As you can see, the separation of the groups are quite obvious.

Last but not least, the most important part of our research is to give recommendations to each group. After we clustered our data, we further dig into each group to identify behavior differences.

We found out that for international vacationers, they tend to like popular and luxury hotels and they are more inclined to buy packages. Therefore, when they are searching for hotels, we recommend Expedia to list high star rating hotels at the top of the search page and devote greater effort to promote package deals to this type of users.

Whereas for business professionals, we found they are more inclined to book branded hotels close to city center. Therefore, it is the best to cater their needs by prioritize downtown branded hotels in their search results.

For the last group, family travelers, since they travel in large groups, usually more than 3 people, we think it is better to recommend them large suites. Since they usually bring their children with them, Expedia can display family activity commercial on their searching page.

In conclusion, we are convinced that our customized recommendations for each user group could help Expedia to increase booking rate among users in United States.