
Travel Agency Case Study

Briana Palencia, Edda Phillips, Jigna Chaudhary, Pratik More, Srikanth Chandesure
BANA 6610

[Video Presentation Link](#)

CONTENTS:

Executive Summary (pg. 1)

Documentation (pg. 3)

Appendix (pg. 8)

To: Travel Agency Management
From: Briana Palencia, Edda Phillips, Jigna Chaudhary, Pratik More, Srikanth Chandesure
Subject: Analysis (2010 - 2022) to Improve Sales Post-COVID
Date: September 14, 2024

This report summarizes patterns in the travel agency's data that may help boost sales after the drop in travel after COVID-19. This report includes recommendations and promotional strategies to enhance sales performance in the future.

EXECUTIVE SUMMARY

Major Findings

- The COVID-19 pandemic caused a 77% decline in yearly profits from 2019 to 2020. After 2020, profits improved by 32%, but 2022 profit is still almost 70% lower than pre-COVID (Exhibit 1). Destination packages displayed a similar trend in decline and growth (Exhibit 2).
- Trends in destination packages varied before and after COVID. Before COVID, Ponoi Russia was the most popular package, but dropped by 74.2% after 2020 (Exhibit 5). The US FIT (United States Flexible Independent Traveler) became the most popular package after 2020, gaining 38.5% of the packages chosen by clients (Exhibit 3).
- Repeat clients are the primary booking source, contributing to 60.8% of the total profit (Exhibit 4).
- Company referrals from 2010 to 2013 may have increased the clientele pool. After company referrals ceased in 2013, repeat clients became the dominant source of bookings, increasing to 853 clients in 2014 compared to 52 clients in 2011, a 1,540% increase (Exhibit 5).
- The photo safaris in the African category generated the highest profit before and after COVID-19, with an average profit of \$5,621 before 2020 and \$6,628 after 2020 (Exhibit 6).
- Clients tend to book almost the same number of vacation days post-COVID as in previous years. The average trip length booked after COVID-19 dropped by only 9% (Exhibit 7).
- Same-day trips are popular among groups of two or fewer (Exhibit 8).
- The heatmap shows that larger groups tend to bring in higher profits. For example, a group size of 19 traveling on an "Intl. Cruise" resulted in the highest profit of \$38,592, while smaller group sizes, such as 1 to 3, mostly generated lower profits, with values ranging from \$907 to \$3,231. Similarly, for the "S. Africa FIT" package, a group of 16 brought in a profit of \$30,465, which is much higher than smaller groups traveling to the same destination (Exhibit 9).

Recommendations for Actions

- Start up the company referrals program again. As shown in the past, company referrals may have boosted the highest booking source (repeat clients), which is the agency's primary source of profit.

- Since the average group size is two, obtain clients' birthdays and anniversaries to offer special promotions for couples during those months.
- Clients are traveling overseas less than they did pre-COVID. Promote new post-pandemic travel preferences by offering special deals with more packages in the United States.
- Focus marketing efforts on repeat clients and less on broader, more generalized Internet and mailing methods.
- To boost profit, increase prices for popular trips, such as US FIT, and decrease the price for the most profitable - yet not as popular - package, African safari.
- Target same-day trip destinations with extended-stay packages.
- Develop specialized packages for specific group sizes to optimize profit.
- Flexible Booking: Provide certainty by allowing for flexible cancellations and bookings.
- Emerging Destinations: Make the most of well-liked post-COVID locations through advertising, packages, and regional alliances.
- Enhance loyalty programs, provide personalized suggestions, exclusive offers, and interesting content to retain customers.

Analytical Overview

Since businesses aim to maximize profit rather than revenue, our analysis focuses on each variable's effects on profit. We focused on the destination packages rather than the destination countries due to the inconsistencies and unreliability in the countries' data. After preparing the data for analysis (a detailed report is annotated in our coding file), we conducted univariate and bivariate analyses. Summary statistics was performed on destination package, trip category, trip duration, group size, booking source, revenue and profit to better understand our client base and assess financial performance. Time-based analyses were then conducted to examine trends from 2010 to 2022, aiming to understand changes in total profits over the years and assess the popularity of destination packages before and after the COVID-19 pandemic (Exhibit 1 and 2). Pie charts were used to compare the top five destination packages before and after COVID-19 and to determine which booking sources were profitable (Exhibit 3 and 4). Correlation analysis was also performed between booking referrals and the number of bookings to identify which referral source resulted in the most bookings (Exhibit 5). Additionally, we analyzed the profitability of each trip type to determine which was most profitable before and after COVID-19 (Exhibit 6). Duration of trips before and after COVID-19 were also compared using categorical data visualizations to identify if there were any differences in trip duration (Exhibit 7). Furthermore, we explored the relationship between group size and same-day trips (Exhibit 8). Finally, a heat map analysis was conducted to determine if target pricing strategies could be applied to group sizes across the top 10 destination packages based on profit generation (Exhibit 9).

DOCUMENTATION PAGES

Exhibit 1: Line plot of profit from sales from 2010-2022

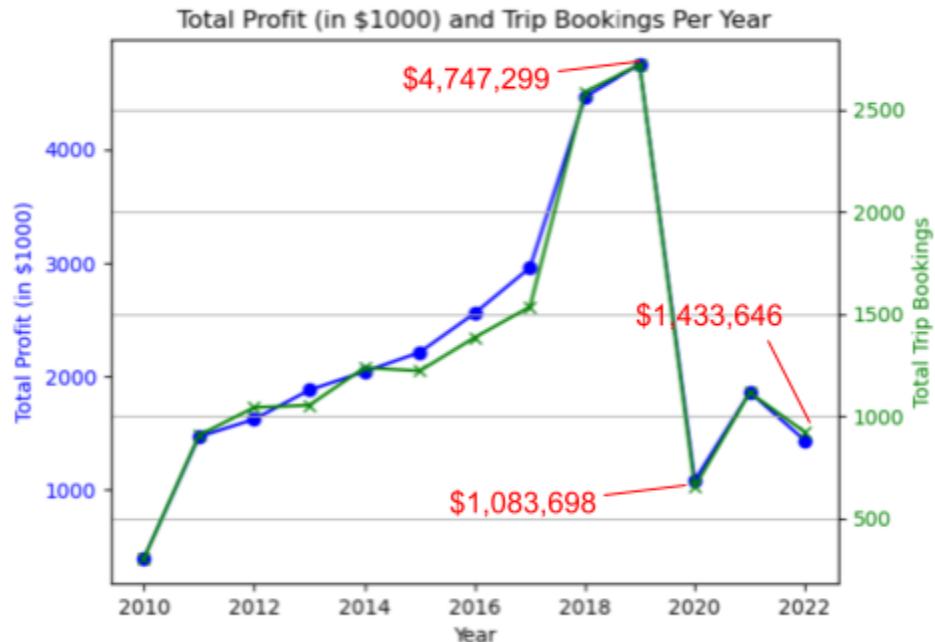


Exhibit 2: Line plot of trends in top 10 destination packages from 2010-2022

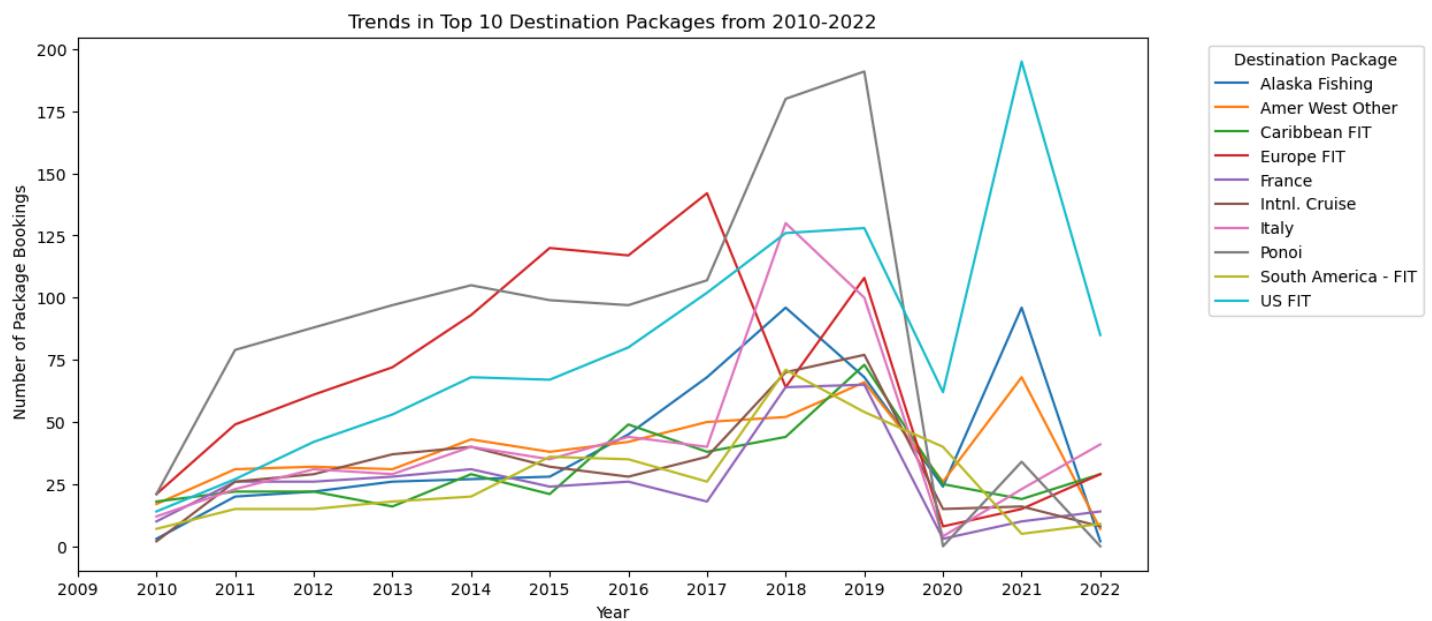


Exhibit 3: Pie charts comparing the 5 top destination packages before and after COVID-19

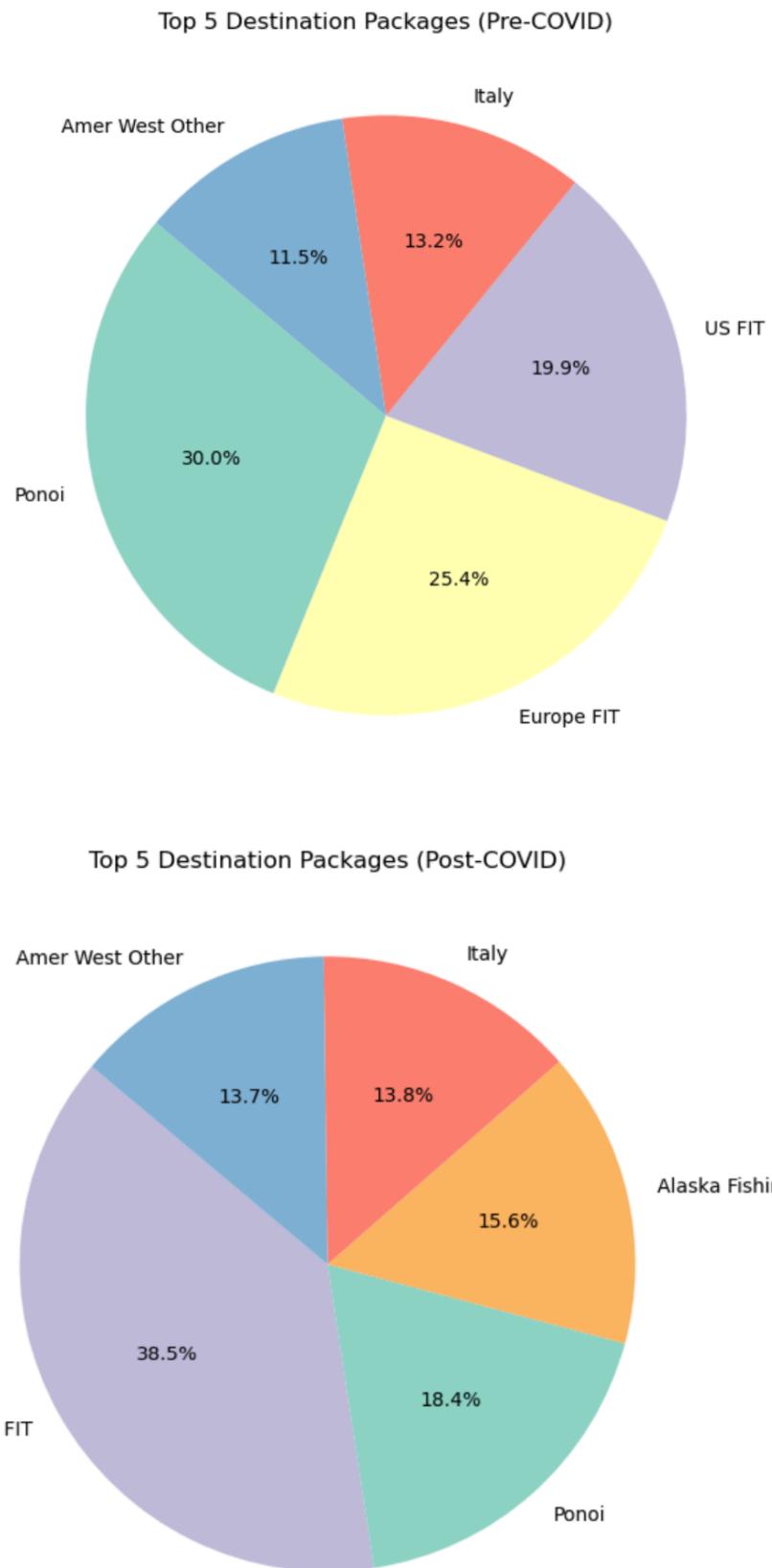


Exhibit 4: Pie chart of most profitable booking sources

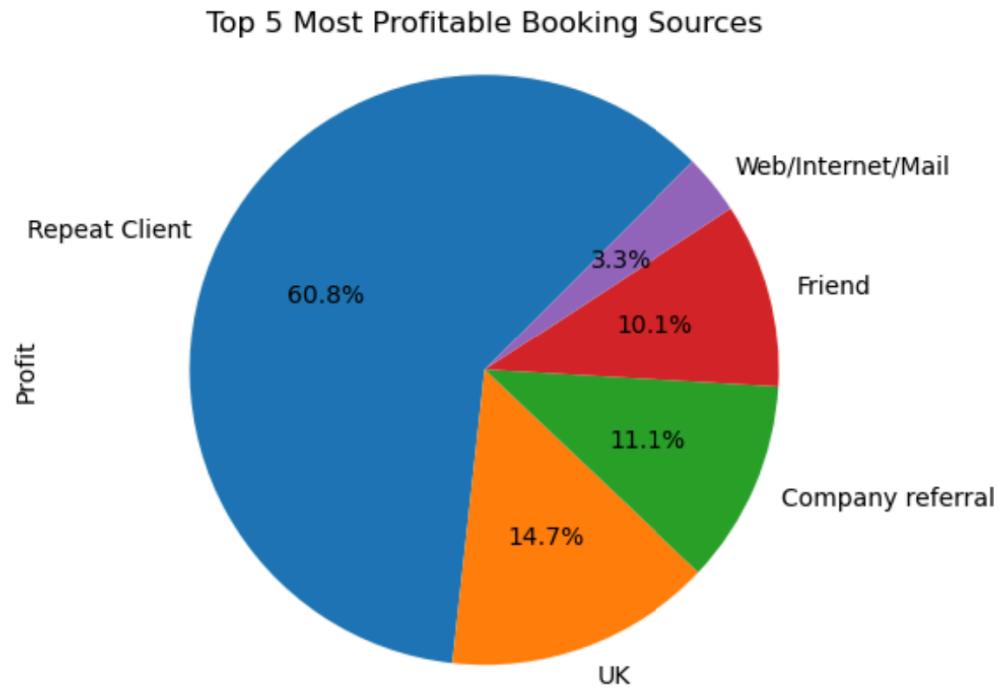


Exhibit 5: Trends in booking in the top three booking sources by the number of trips booked yearly

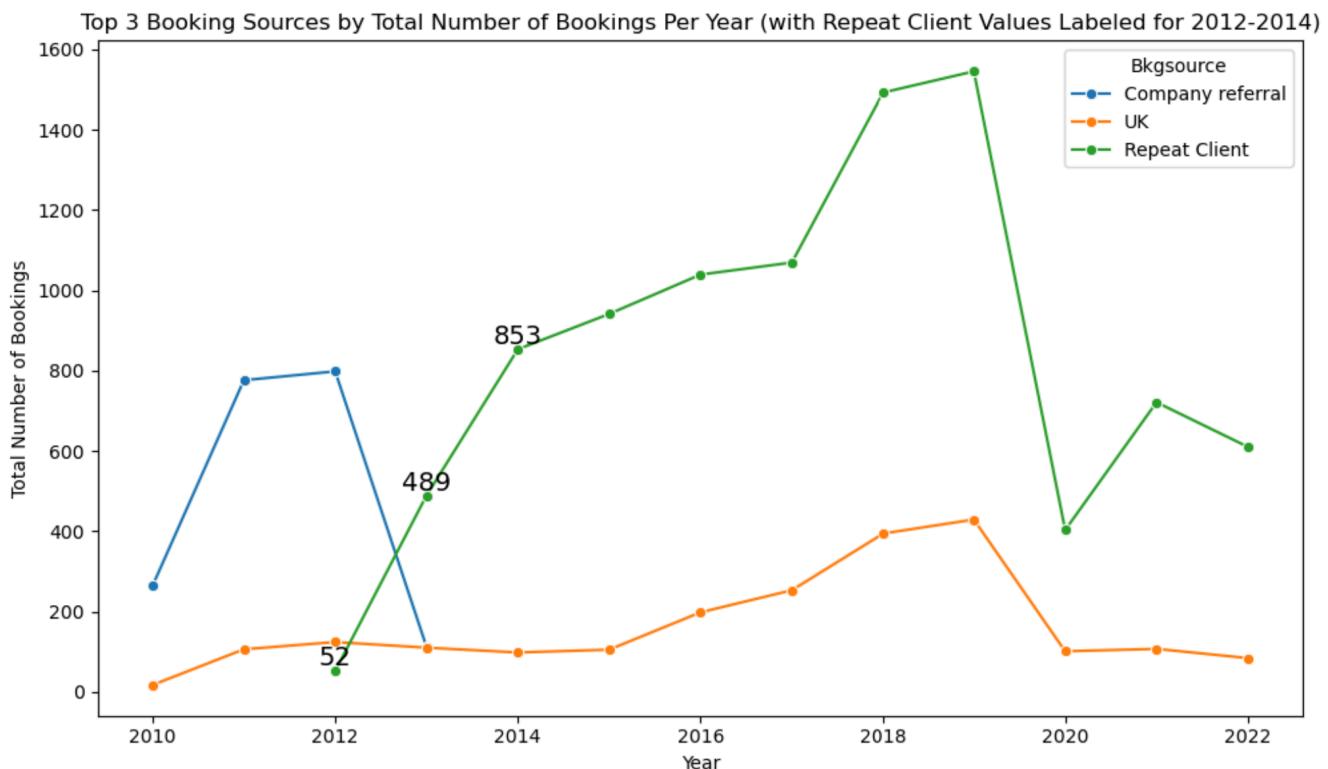


Exhibit 6: Bar chart of the most profitable type of trip for bookings before and after 2020

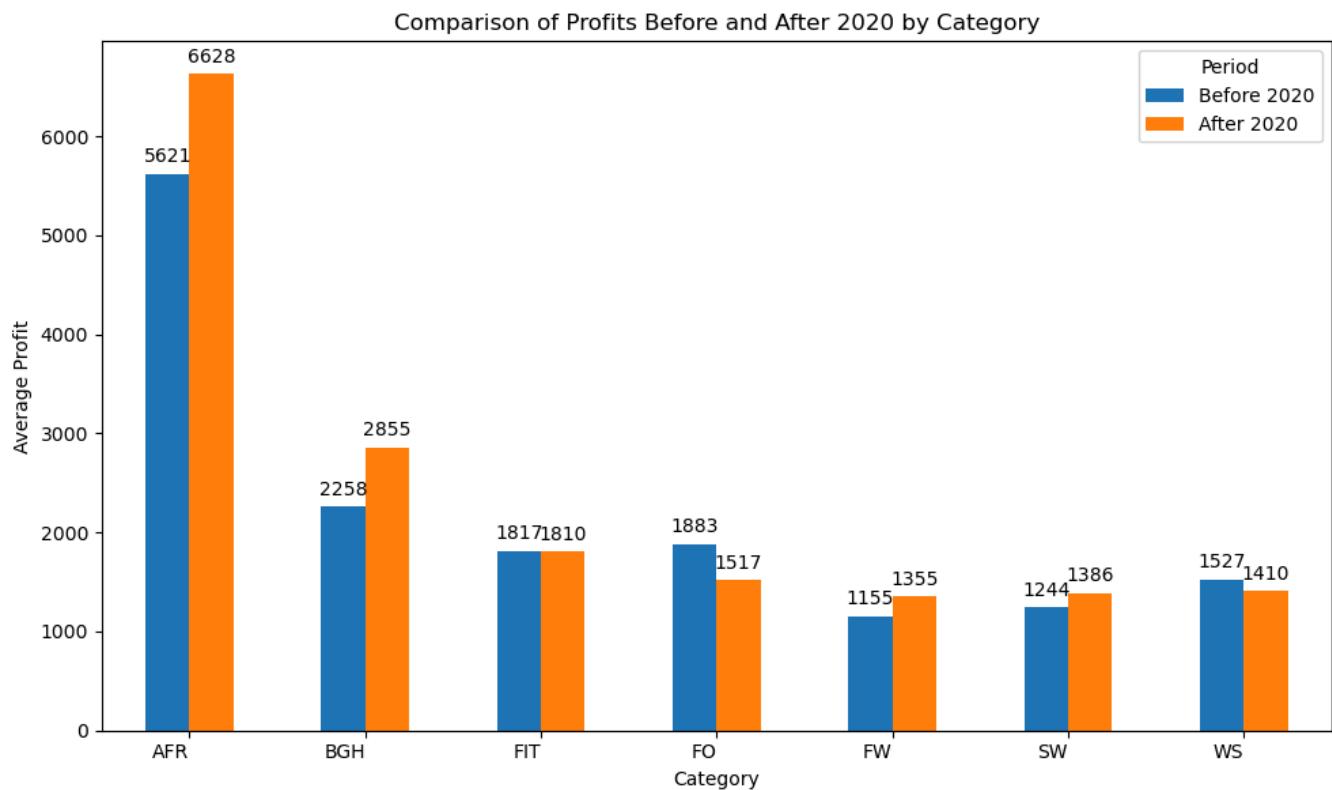


Exhibit 7: Bar graphs comparing the duration of trips pre- and post-COVID from 2010-2022 by month

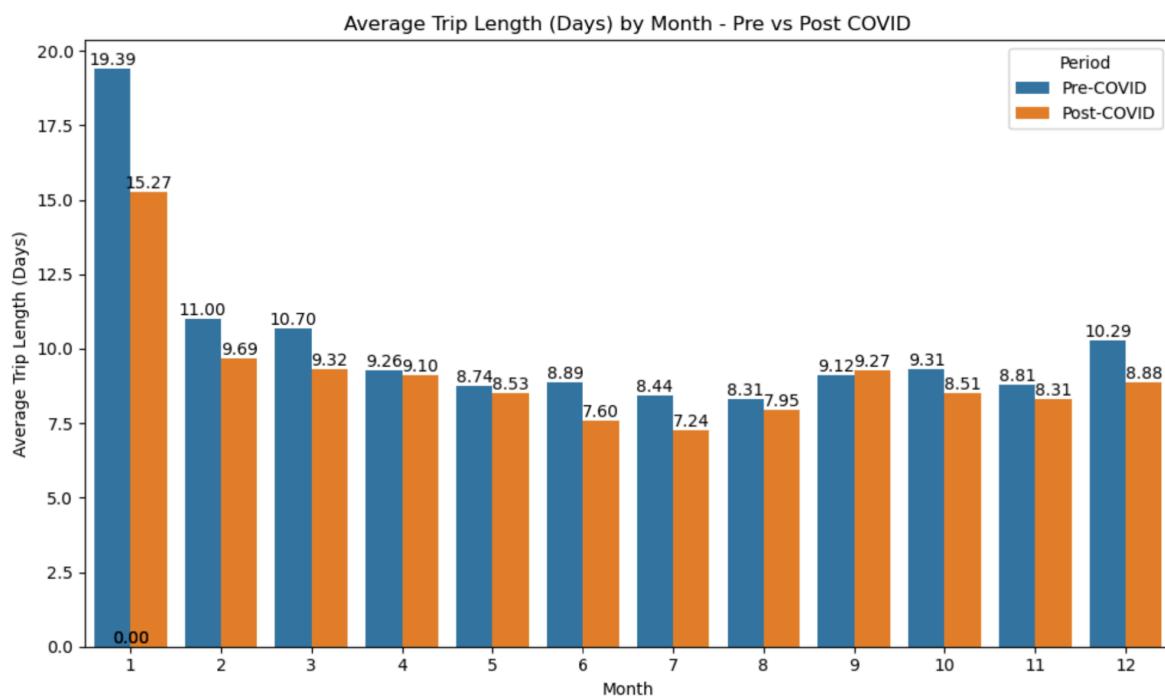


Exhibit 8: Distribution of group size and the number of trips

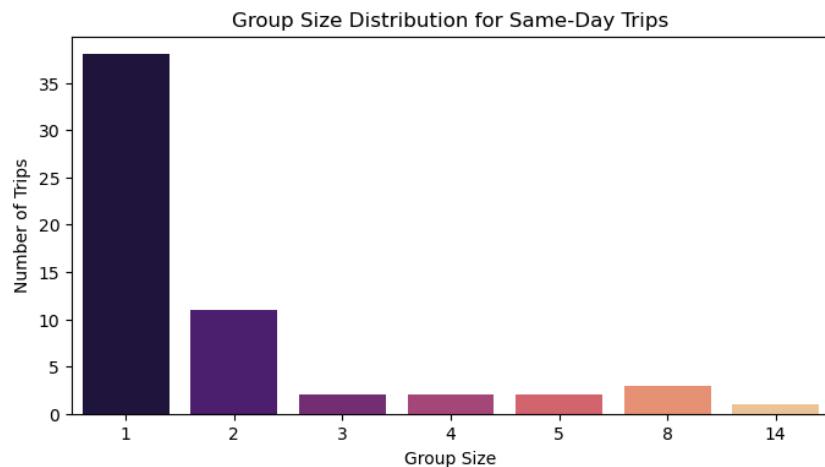
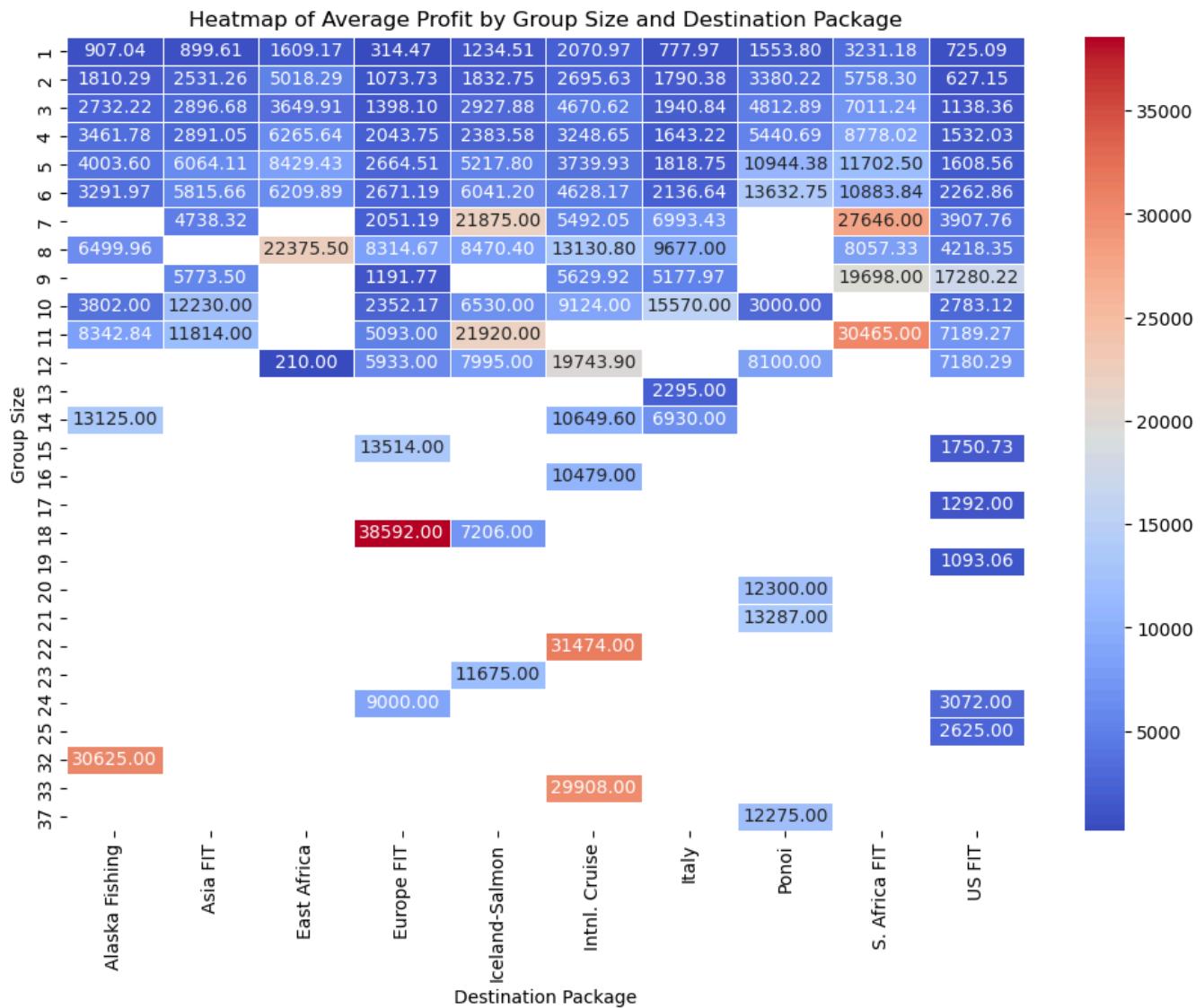


Exhibit 9: Heatmap of average profit by group size and destination package



APPENDIX

Appendix A: Data cleaning of the travel agency's data

Objective: Analyze past booking patterns and sales trends to provide insight and strategies for enhancing sales.

In-House Experts: Clarified some terms with Professor Du and Google

Domain Knowledge: Technofunk: Components of Tourism Industry (useful website)

Population of Interest: Travelers from this agency's records from 2009 until 2022. Representative Sample.

Relevant Independent Variables: Days of Trip, Month, Destination Packages, Category, Groups Size, Booking Source

Data Cleansing Notes

We imported traveldata.csv to excel and cleaned it there.

- We verified each column had the same number of rows.
- We hid the original columns clientID since it wasn't needed.
- We trimmed white spaces
- For the null/NaN/UK cells we replaced with "Other."
- We checked the consistency of the data types, making all categorical variables string and numerical variables either float_64 or int.
- Checked for consistent entries
- Replaced US States listed as countries to US
- Clarified column titles and capitalized
- Checked and fixed Trip IDs with multiple destinations
- Checked and fixed the variations of spelling in each destination_package

We imported the Excel-cleaned data into JupyterLab and verified the data was still clean.

Other things to note:

- Since businesses should focus on maximizing profit, not revenue, we focused on profit as the key dependent variable
- There were many errors in the Destination_Countries column (i.e. duplicate names, incorrect countries, states listed as their own country, etc.) Because of this, we focused our analysis on Destination Packages. We verified unique package names.
- A booking source "UK" needs to be clarified. Is this "Unknown" or "United Kingdom" perhaps? We left it as "UK."
- We created the column "Trip_Days" by subtracting the "Date_of_trip_to" from "Date_of_trip_to" column.
- We corrected a date that initially listed the trip as -21. Suspecting this might be a typo, we adjusted the trip date to match the date listed in the "trip to" column.

```
: import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats
from scipy.stats import describe
```

```
df=pd.read_csv('travel_data_3.csv')
df.head()
```



bkgref-test	Trip_ID	Destination_Package	Destination_Country	Category	Date_of_trip_from	Date_of_trip_to	Trip_days	Year	Month_No	Month	bkgsource	Revenue	Profit	Groupsize	
0	B16-42027	805	Galapagos-FIT	Ecuador	FIT	4/7/2017	4/16/2017	9	2017	4	Apr	Repeat Client	20910.0	2322.0	3
1	B11-25115	211	Deep Water Cay	Bahamas	SW	4/13/2011	4/17/2011	4	2011	4	Apr	Company referral	3295.0	659.0	1
2	B11-27267	211	Deep Water Cay	Bahamas	SW	4/23/2012	4/27/2012	4	2012	4	Apr	Company referral	3900.0	696.9	1
3	B11-25462	211	Deep Water Cay	Bahamas	SW	4/28/2011	5/2/2011	4	2011	4	Apr	Company referral	2805.0	503.0	1
4	B14-36987	235	North Riding Point	Bahamas	SW	4/5/2015	4/11/2015	6	2015	4	Apr	Repeat Client	9600.0	1920.0	2

```
df = df.drop('Bkgsource', axis=1)
df.head()
```

bkgref-test	Trip_ID	Destination_Package	Destination_Country	Category	Date_of_trip_from	Date_of_trip_to	Trip_days	Year	Month_No	Month	Revenue	Profit	Groupsize	Bkgsource	
0	B16-42027	805	Galapagos-FIT	Ecuador	FIT	4/7/2017	4/16/2017	9	2017	4	Apr	20910.0	2322.0	3	Repeat Client
1	B11-25115	211	Deep Water Cay	Bahamas	SW	4/13/2011	4/17/2011	4	2011	4	Apr	3295.0	659.0	1	Company referral
2	B11-27267	211	Deep Water Cay	Bahamas	SW	4/23/2012	4/27/2012	4	2012	4	Apr	3900.0	696.9	1	Company referral
3	B11-25462	211	Deep Water Cay	Bahamas	SW	4/28/2011	5/2/2011	4	2011	4	Apr	2805.0	503.0	1	Company referral
4	B14-36987	235	North Riding Point	Bahamas	SW	4/5/2015	4/11/2015	6	2015	4	Apr	9600.0	1920.0	2	Repeat Client

Appendix B: Summary statistics of package, trip category, duration of trip, month, group size, booking referral, revenue and profit

```
select_columns = ['Destination_Package', 'Category', 'Trip_days', 'Month', 'Groupsize', 'Bkgsource']
mode_values = df[select_columns].mode().iloc[0]
mode_counts = df[select_columns].apply(lambda col: (col == mode_values[col.name]).sum())
total_rows = len(df)
mode_percentages = (mode_counts / total_rows) * 100

modes_and_counts = pd.DataFrame({
    'Mode': mode_values,
    'Count': mode_counts,
    'Percentage (%)': mode_percentages
})

from IPython.display import display, HTML
display(HTML("<h3>Most Popular Categories</h3>"))
display(modes_and_counts)
```

Most Popular Categories

	Mode	Count	Percentage (%)
Destination_Package	Ponoi	1098	6.582339
Category	FIT	6073	36.406690
Trip_days	7	3798	22.768419
Month	Jun	1944	11.653978
Groupsize	1	7691	46.106349
Bkgsource	Repeat Client	9214	55.236497

```
metric_columns = ['Trip_days', 'Revenue', 'Profit', 'Groupsize']
df[metric_columns].describe()
```

	Trip_days	Revenue	Profit	Groupsize
count	16681.000000	16681.000000	16681.000000	16681.000000
mean	9.692045	12049.907307	1721.492028	1.970146
std	17.954547	18780.003225	2519.972780	1.764510
min	0.000000	0.000000	-400.000000	1.000000
25%	6.000000	4050.000000	590.000000	1.000000
50%	7.000000	7190.000000	1072.000000	2.000000
75%	10.000000	13666.000000	2000.000000	2.000000
max	386.000000	630000.000000	63000.000000	37.000000

Appendix C: Python code for Exhibit 1 – Line plot of profit from sales (2010-2022)

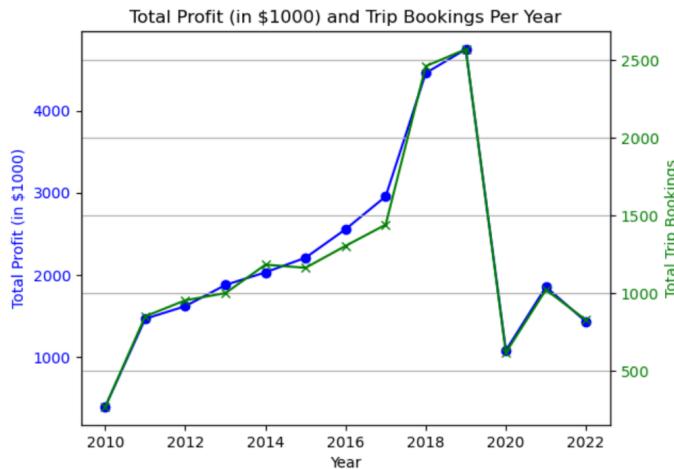
```
[522]: total_profit_per_year = df_noexpref.groupby('Year')['Profit'].sum()
total_profit_per_year_in_thousands = total_profit_per_year / 1000
total_trips_per_year = df_noexpref.groupby('Year').size()

fig, ax1 = plt.subplots()

ax1.set_xlabel('Year')
ax1.set_ylabel('Total Profit (in $1000)', color='blue')
ax1.plot(total_profit_per_year_in_thousands.index, total_profit_per_year_in_thousands.values, marker='o', color='blue', label='Total Profit (in $1000)')
ax1.tick_params(axis='y', labelcolor='blue')

ax2 = ax1.twinx()
ax2.set_ylabel('Total Trip Bookings', color='green')
ax2.plot(total_trips_per_year.index, total_trips_per_year.values, marker='x', color='green', label='Total Trip Bookings')
ax2.tick_params(axis='y', labelcolor='green')
plt.title('Total Profit (in $1000) and Trip Bookings Per Year')

plt.grid(True)
plt.show()
```



Appendix D: Python code for analysis of total profit in 2019, 2020, and 2022, and percentage of profit loss between these years

```
[558]: total_profit_2019 = df[df['Year'] == 2019]['Profit'].sum()
total_profit_2020 = df[df['Year'] == 2020]['Profit'].sum()
total_profit_2022 = df[df['Year'] == 2022]['Profit'].sum()

if total_profit_2019 != 0:
    percentage_loss_1920 = ((total_profit_2019 - total_profit_2020) / total_profit_2019) * 100
else:
    percentage_loss = 0

if total_profit_2020 != 0:
    percentage_gain_2220 = ((total_profit_2022 - total_profit_2020) / total_profit_2020) * 100
else:
    percentage_gain_2220 = 0

if total_profit_2019 != 0:
    percentage_loss_1922 = ((total_profit_2019 - total_profit_2022) / total_profit_2019) * 100
else:
    percentage_loss = 0

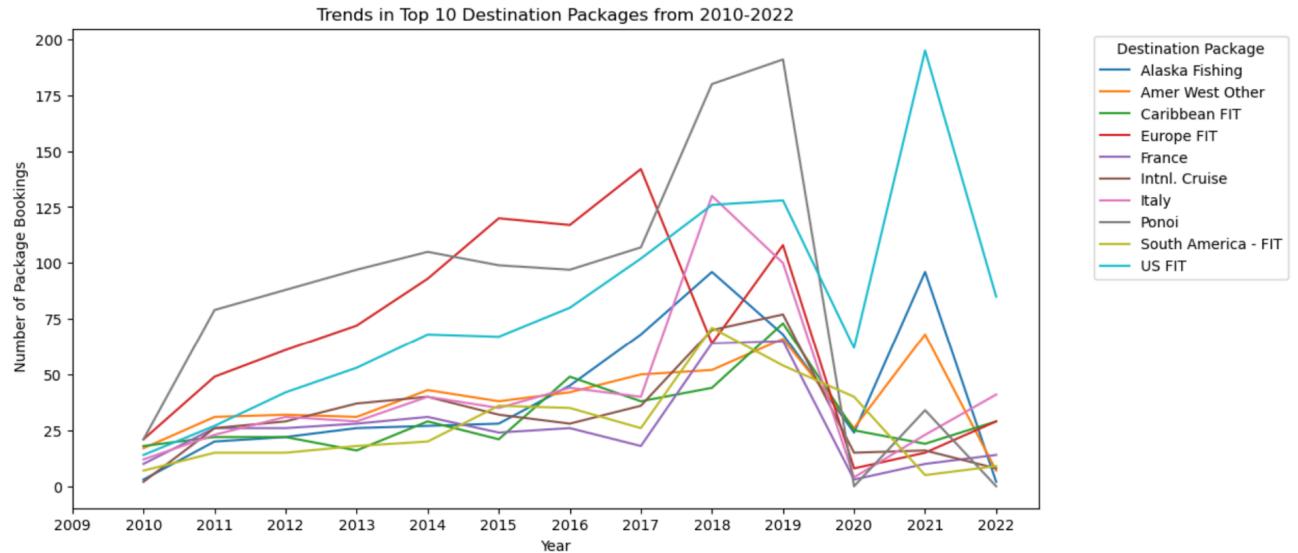
print(f"The percentage of profits lost from 2019 to 2020 is {percentage_loss_1920:.2f}%.")
print(f"The percentage of profits lost from 2019 to 2022 is {percentage_loss_1922:.2f}%.")
print(f"The percentage of profits gained from 2020 to 2022 is {percentage_gain_2220:.2f}%.")
print(f'The total profit in 2019 is {total_profit_2019}')
print(f'The total profit in 2020 is {total_profit_2020}')
print(f'The total profit in 2022 is {total_profit_2022}')
```

The percentage of profits lost from 2019 to 2020 is 77.17%.
The percentage of profits lost from 2019 to 2022 is 69.80%.
The percentage of profits gained from 2020 to 2022 is 32.29%.
The total profit in 2019 is 4747299.284
The total profit in 2020 is 1083698.260000002
The total profit in 2022 is 1433646.989999998

Appendix E: Python code for Exhibit 2 – Line plot of trends in top 10 destination packages (2010-2022)

```
[484]: top_10_packages = df['Destination_Package'].value_counts().nlargest(10).index
df_top_10 = df[df['Destination_Package'].isin(top_10_packages)]

top_packages_over_time = df_top_10.groupby(['Year', 'Destination_Package']).size().unstack().fillna(0)
top_packages_over_time.plot(figsize=(12, 6), kind='line')
plt.title('Trends in Top 10 Destination Packages from 2010-2022')
plt.xlabel('Year')
plt.ylabel('Number of Package Bookings')
plt.legend(title='Destination Package', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.xticks(ticks= list(range(2009,2023)))
plt.show()
```



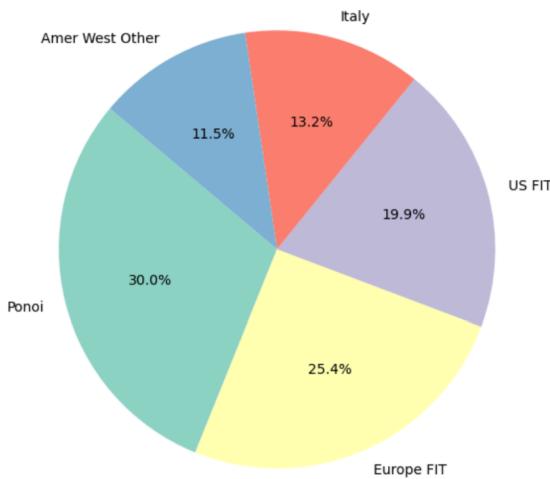
All the top 10 destination packages faced a decline in bookings during 2020 (when COVID-19 happened). Number of bookings for the packages increased again in 2021 for Ponoi (grey), Amer West Other (orange), Alaska Fishing (dark blue) and US FIT (light blue). In general, there was an increase in bookings for the top 10 destination over the years, until 2020.

Appendix F: Python code for Exhibit 3 – Pie charts comparing the top 5 destination packages before and after COVID-19

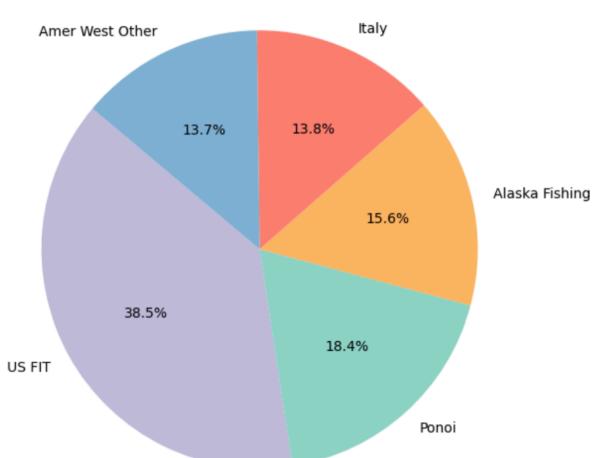
```
[443]: df['Year'] = pd.to_datetime(df['Date_of_trip_from']).dt.year
pre_covid_data = df[df['Year'] < 2019]
post_covid_data = df[df['Year'] >= 2019]
pre_covid_trips = pre_covid_data.groupby('Destination_Package').size().reset_index(name='Trips_PreCOVID')
post_covid_trips = post_covid_data.groupby('Destination_Package').size().reset_index(name='Trips_PostCOVID')
top_5_pre_covid = pre_covid_trips.nlargest(5, 'Trips_PreCOVID')
top_5_post_covid = post_covid_trips.nlargest(5, 'Trips_PostCOVID')
all_top_destinations = pd.concat([top_5_pre_covid, top_5_post_covid])['Destination_Package'].unique()

colors = plt.get_cmap('Set3')(range(len(all_top_destinations)))
color_map = {destination: colors[i] for i, destination in enumerate(all_top_destinations)}
pre_covid_colors = [color_map[dest] for dest in top_5_pre_covid['Destination_Package']]
post_covid_colors = [color_map[dest] for dest in top_5_post_covid['Destination_Package']]
plt.figure(figsize=(14, 6))
plt.subplot(1, 2, 1)
plt.pie(top_5_pre_covid['Trips_PreCOVID'], labels=top_5_pre_covid['Destination_Package'],
        autopct='%1.1f%%', startangle=140, colors=pre_covid_colors)
plt.title('Top 5 Destination Packages (Pre-COVID)')
plt.subplot(1, 2, 2)
plt.pie(top_5_post_covid['Trips_PostCOVID'], labels=top_5_post_covid['Destination_Package'],
        autopct='%1.1f%%', startangle=140, colors=post_covid_colors)
plt.title('Top 5 Destination Packages (Post-COVID)')
plt.tight_layout()
plt.show()
```

Top 5 Destination Packages (Pre-COVID)



Top 5 Destination Packages (Post-COVID)



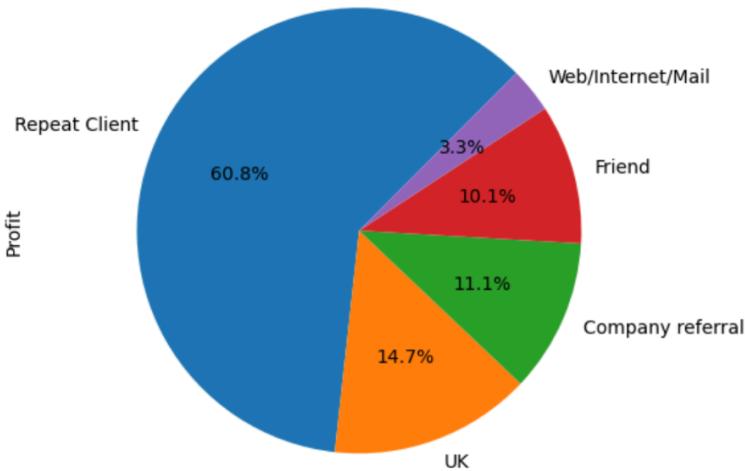
Appendix G: Python code for Exhibit 4 – Pie chart of most profitable booking sources

```
[442]: bkgsource_profit = df_norevprof.groupby('Bkgsource')['Profit'].sum()

top_5_bkgsource_profit = bkgsource_profit.sort_values(ascending=False).head(5)
top_5_bkgsource_profit.plot(
    kind='pie', autopct='%1.1f%%', startangle=45)

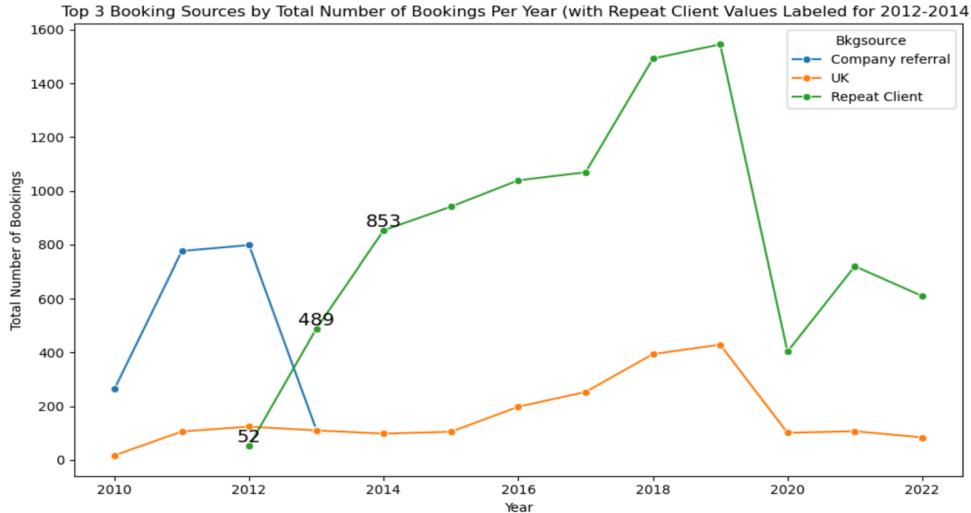
plt.title('Top 5 Most Profitable Booking Sources')
plt.axis('equal')
plt.show()
```

Top 5 Most Profitable Booking Sources



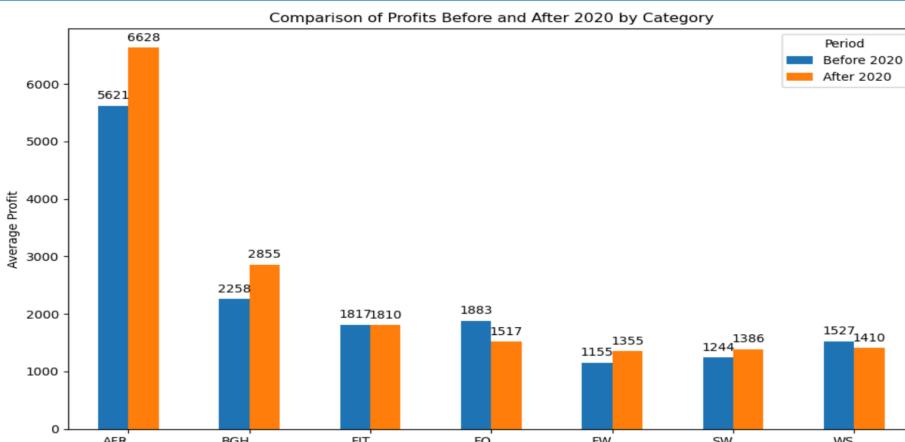
Appendix H: Python code for Exhibit 5 – Trends in bookings from the top three booking sources by number of trips booked yearly

```
[486]: df['IsRepeatClient'] = df['Bkgsource'].apply(lambda x: 1 if x == 'Repeat Client' else 0)
bookings_per_year = df.groupby(['Year', 'Bkgsource']).size().reset_index(name='Total Bookings')
top_3_sources = bookings_per_year.groupby('Bkgsource')['Total Bookings'].sum().nlargest(3).index
top_3_data = bookings_per_year[bookings_per_year['Bkgsource'].isin(top_3_sources)]
repeat_clients = df[(df['Bkgsource'] == 'Repeat Client') & (df['Year'].between(2012, 2014))]
repeat_clients_per_year = repeat_clients.groupby('Year').size().reset_index(name='Repeat Bookings')
plt.figure(figsize=(10, 6))
sns.lineplot(x='Year', y='Total Bookings', hue='Bkgsource', data=top_3_data, marker='o')
for _, row in repeat_clients_per_year.iterrows():
    plt.text(row['Year'], row['Repeat Bookings'], f'{row["Repeat Bookings"]}', color='black', size=14, ha='center', va='bottom')
plt.title('Top 3 Booking Sources by Total Number of Bookings Per Year (with Repeat Client Values Labeled for 2012-2014)')
plt.xlabel('Year')
plt.ylabel('Total Number of Bookings')
plt.tight_layout()
plt.show()
```

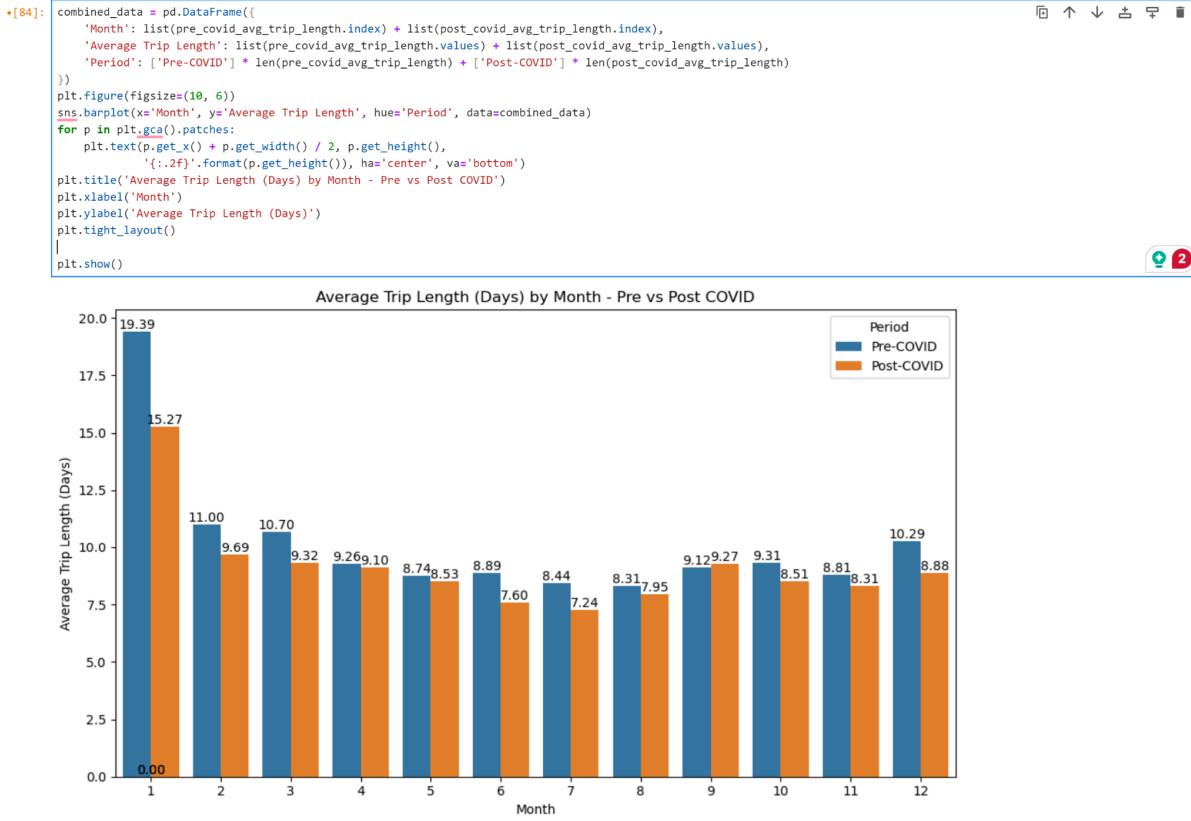


Appendix I: Python code for Exhibit 6 – Bar chart of the most profitable trip types before and after 2020

```
[df['Period'] = df['Year'].apply(lambda x: 'Before 2020' if x <= 2020 else 'After 2020')
avg_profits = df.groupby(['Category', 'Period'])['Profit'].mean().reset_index()
pivot_table = avg_profits.pivot(index='Category', columns='Period', values='Profit').fillna(np.nan)
period_order = ['Before 2020', 'After 2020']
pivot_table = pivot_table[period_order]
category_profit_bar = pivot_table.plot(kind='bar', figsize=(10, 6))
for p in category_profit_bar.patches:
    height = p.get_height()
    if not np.isnan(height):
        category_profit_bar.annotate(format(height, '.0f'),
            (p.get_x() + p.get_width() / 2., height),
            ha='center', va='center', xytext=(0, 0), textcoords='offset points', fontsize=10)
plt.title('Comparison of Profits Before and After 2020 by Category')
plt.xlabel('Category')
plt.ylabel('Average Profit')
plt.xticks(rotation=0, ha='right')
plt.legend(title='Period')
plt.tight_layout()
plt.show()
```



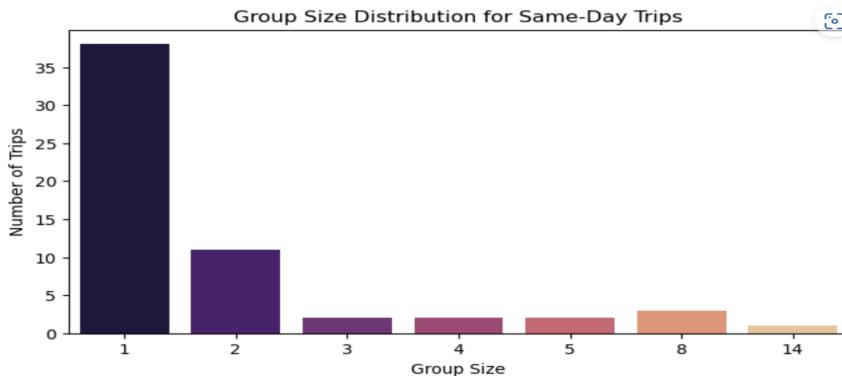
Appendix J: Python code for Exhibit 7 – Bar graphs comparing trip duration pre- and post-COVID (2010-2022) by month



Appendix K: Python code for Exhibit 8 – Distribution of group size and the number of trips

```
zero_trip_period = df[df['Trip_days'] == 0]

# Count the number of zero-period trips and calculate the percentage
zero_trip_count = len(zero_trip_period)
total_count = len(df)
percentage_zero_trip = (zero_trip_count / total_count) * 100
destinations_zero_trip = zero_trip_period['Destination_Country'].value_counts()
groupsize_zero_trip = zero_trip_period['Groupsize'].value_counts()
plt.figure(figsize=(8,4))
sns.barplot(x=groupsize_zero_trip.index, y=groupsize_zero_trip.values, palette='magma')
plt.title('Group Size Distribution for Same-Day Trips')
plt.ylabel('Number of Trips')
plt.xlabel('Group Size')
plt.show()
```



Appendix L: Python code for Exhibit 9 – Heatmap of average profit by group size and destination package

```
[308]: import seaborn as sns
import matplotlib.pyplot as plt

pivot_table = group_recommendation_top.pivot_table(index='groupsize', columns='Destination_Country', values='Profit', aggfunc='mean')
plt.figure(figsize=(12, 8))
sns.heatmap(pivot_table, annot=True, cmap='coolwarm', fmt='.2f', linewidths=.5)
plt.title('Heatmap of Average Profit by Group Size and Destination Country')
plt.xlabel('Destination Country')
plt.ylabel('Group Size')
plt.show()
```

