Instructional Airplane Leasing Program

Overview

This dataset details every accident investigated by the National Transportation Safety Board over the last 40 years. This data details the make, models, weather conditions, flight phases, injuries, and deaths during each accident. The data could potentially help organizations find correlations, and identify conditions of an airplane flight that make it safer.

Key findings:

1. **Cessna** and **Piper** airplanes appear to offer better protection to its passengers during airplane accidents.
2. Accidents that occur during **taxi**, **takeoff**, and **landing** have the **lowest** fatalities rates. Accidents that occur during **maneuvers** and **cruise** phases have the **highest** fatality rates.
3. Flights occurring during the **summer** months tend to have **lower** fatality rates that flights occuring during the **winter** months.

Business Problem

In order to choose safe airplanes for the flight school, we will investigate these business questions:

1. How does the make of the aircraft affect airplane safety?
2. How does the phase of flight affect the survivability of airplane accidents?
3. How do the time of year and weather conditions affect the survival rate an airplane accident?

To measure airplane safety, we will consider the protection offered by the airplane during an accident. We will look at this three different ways

- **Survive:** The fraction of accidents with no fatalities.
- **Fraction Fatal:** The fraction of passengers killed during an accident
- **Fraction Uninjured:** The fraction of passengers uninjured during an accident.

Data Understanding

Intake Data

In this analysis we read in the entire AviationData dataset. For each record, we use the following data:

- Event Date
- Make of Airplane
- Accidents injured, killed, minor injuries, and uninjured passengers in each accident
- Phase of Flight and weather conditions

```python
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         sns.set_theme("paper")

         import scr.data_preparation as dp
         import scr.data_visualizations as dv

         df = pd.read_csv("Data/AviationDataClean.csv", dtype={"Latitude":object, "Lo
                                                               "Report.Status":object
```

```
/var/folders/ym/68nrz1n97wj0gz5413bhpqs80000gn/T/ipykernel_16497/1870390053.p
y:10: DtypeWarning: Columns (32) have mixed types. Specify dtype option on im
port or set low_memory=False.
  df = pd.read_csv("Data/AviationDataClean.csv", dtype={"Latitude":object, "L
ongitude":object,
```

To create a final data feature, a concatenation of the Make and Models, we drop entries missing values and join strings together.

In [2]:
```python
df = df.dropna(subset="Model")
df["Make_Model"] = df[["Make","Model"]].apply(lambda x: "-".join(x), axis=1)
```

Outcome Data

Analysis

Airplane Safety by Make of Aircraft

First, we will identify airplane manufacturers that make airplanes with the best protection during accidents. To do this, we average the statistic survive across all accidents associated with each manufacturer, and graph the results. The graph shows that Cessna, Bell, and Piper aircraft offer the best protection during accidents. There are no fatalities in 80% of accidents involving their aircraft. Mooney and Beech perform 10% worse, with only 70% of accidents with their aircraft involving no fatalities.

In [3]:
```python
fig, ax = plt.subplots(1,2, figsize = (10,4))

sns.barplot(data=df, x='Make', y = 'Survive', ax = ax[0], order=df[["Make","
sns.pointplot(data=df, x='Make', y = 'Survive', order=df[["Make","Survive"]]

fig.savefig("Images/SurviveMake")
```
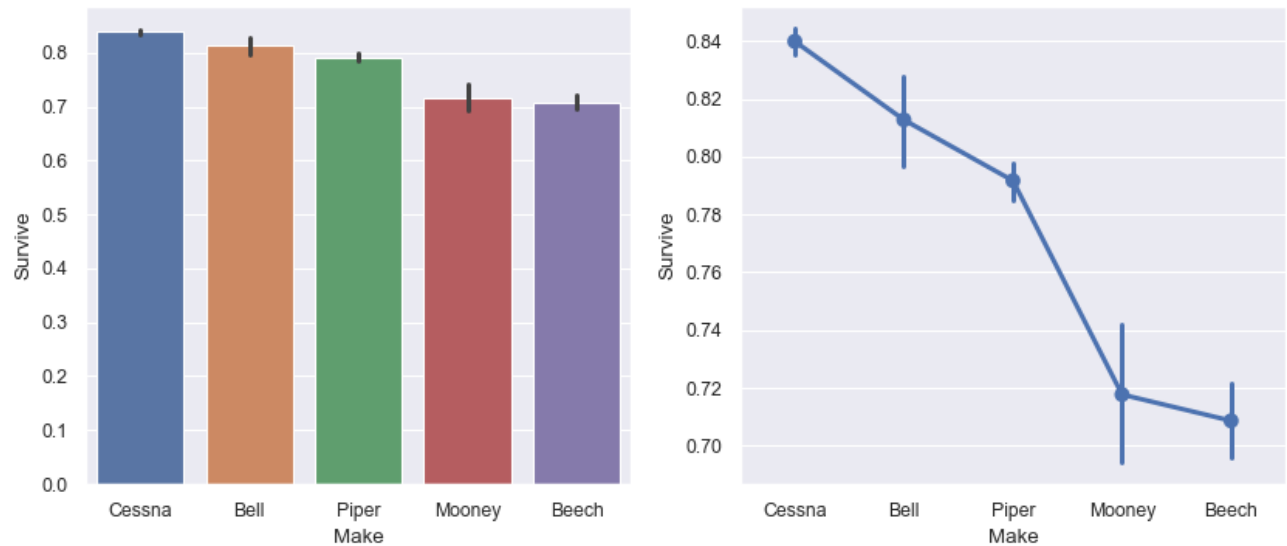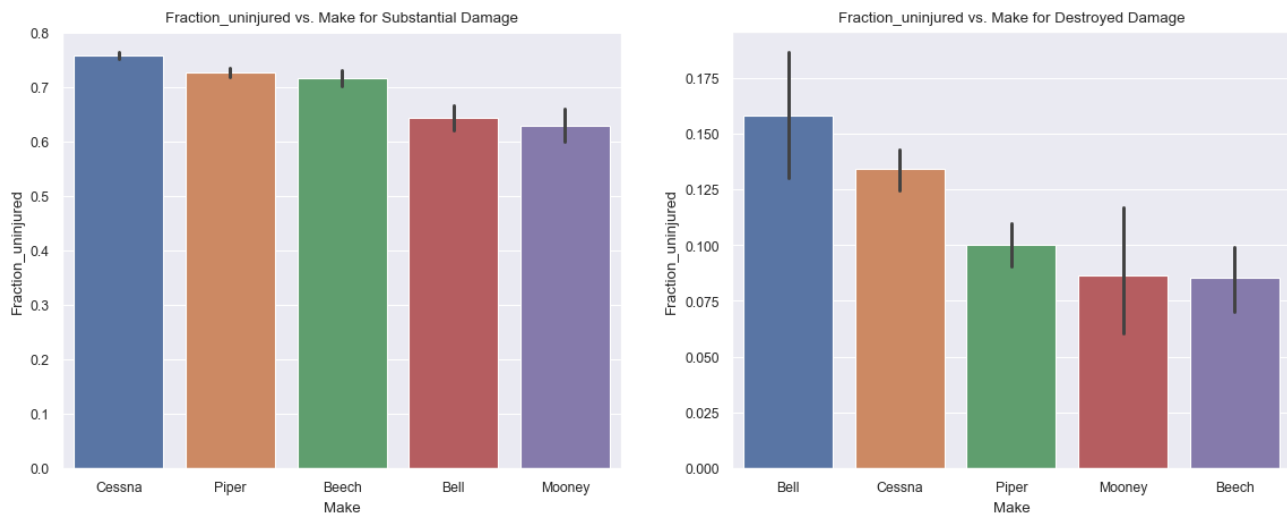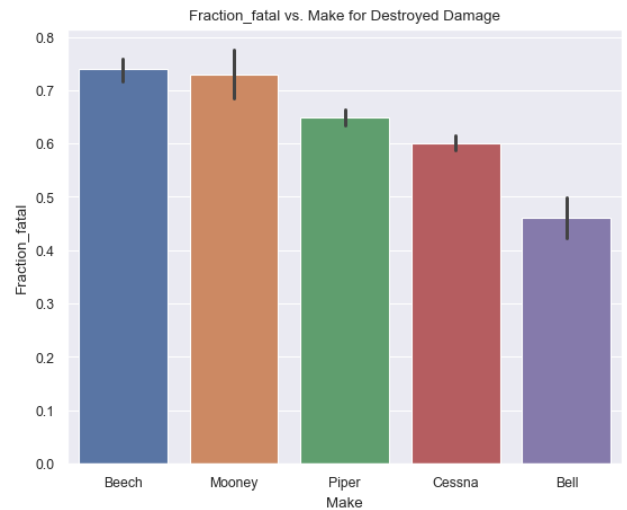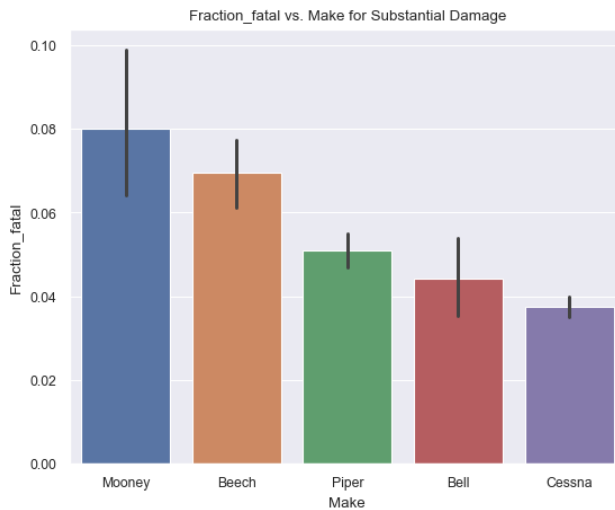
When breaking down accidents further to consider whether aircraft sustain substantial damage or are destroyed, we find that Cessna, Bell, and Piper makes consistently offer the best protection of the five.

In [4]: `dv.performance_make_damage(df, "Fraction_uninjured")`

In [5]: `dv.performance_make_damage(df, "Fraction_fatal")`

Model Safety

Knowing the Cessna, Piper, and Bell offer the best protection during accidents, we look at specific models manufactured by each, and identify which are the safest. Again, we break down the statistics to separately consider cases where aircraft are substantially damaged and destroyed.
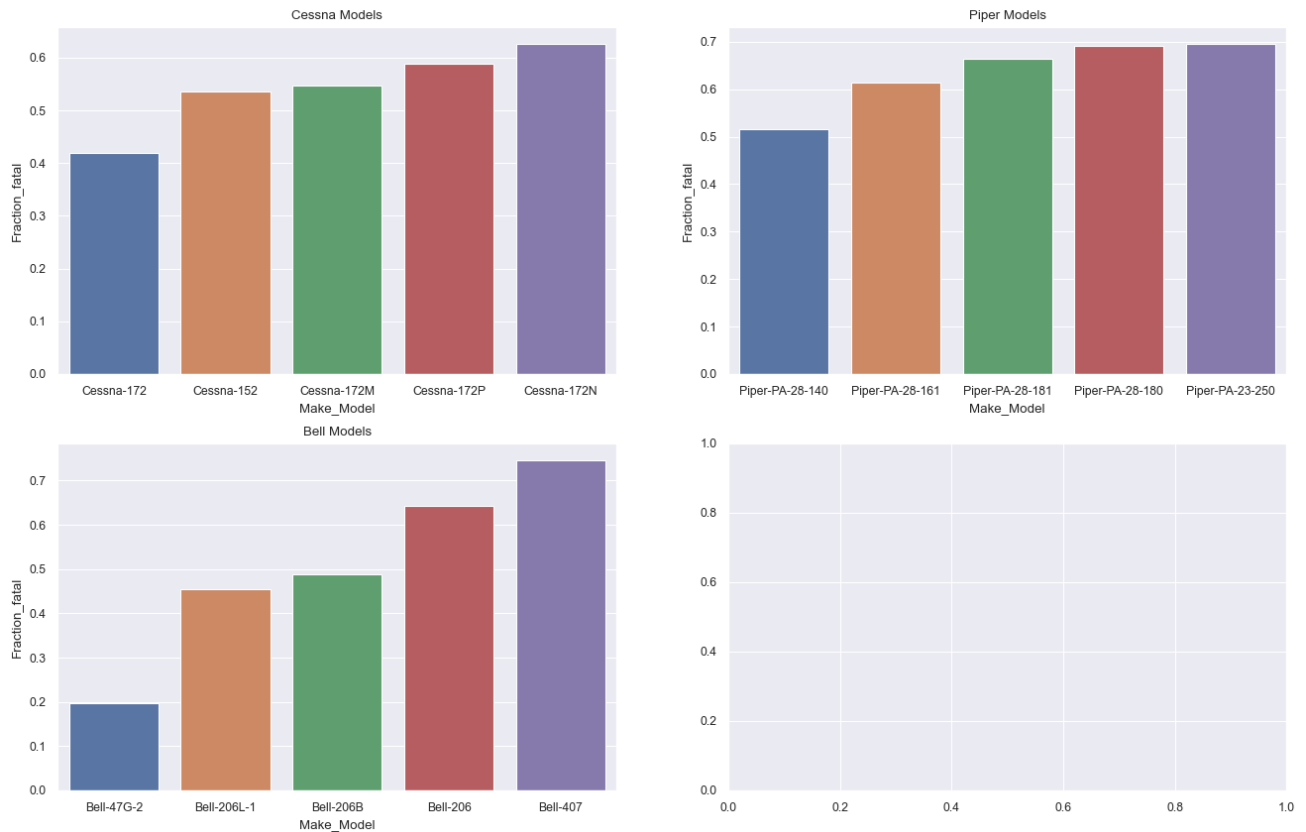
In [6]:
```python
dfD = df[df['Aircraft.damage'] == 'Destroyed']
fig, ax = plt.subplots(2,2, figsize = (16,10))

fig.suptitle('Destroyed per Make-Model in Fatal', fontsize = 18)

dv.hi(dfD, 'Cessna', 'Fraction_fatal', ax[0][0])
dv.hi(dfD, 'Piper', 'Fraction_fatal', ax[0][1])
dv.hi(dfD, 'Bell', 'Fraction_fatal', ax[1][0])

fig.savefig("Images/ModelDestroyedFatal")
```

Destroyed per Make-Model in Fatal


```
In [7]:  dfD = df[df['Aircraft.damage'] == 'Substantial']
         fig, ax = plt.subplots(2,2, figsize = (16,10))

         fig.suptitle('Substantial Damage per Make-Model in Fatal', fontsize = 18)

         dv.hi(dfD, 'Cessna', 'Fraction_fatal', ax[0][0])
         dv.hi(dfD, 'Piper', 'Fraction_fatal', ax[0][1])
         dv.hi(dfD, 'Bell', 'Fraction_fatal', ax[1][0])

         fig.savefig("Images/ModelSubstantialFatal")
```

Substantial Damage per Make-Model in Fatal
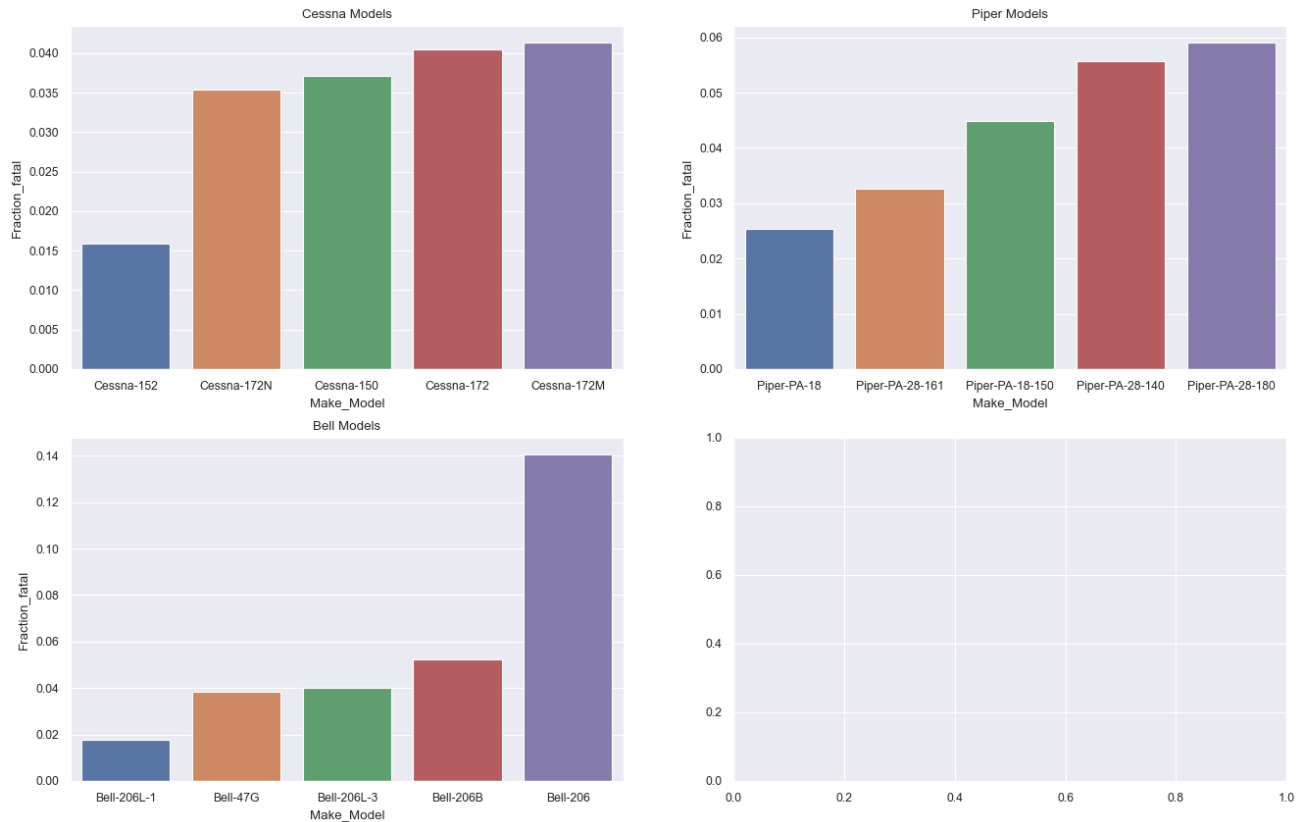
Based on the above analysis, we compare the top 5 models, and find that the Cessna 172, Cessna 152, and Piper PA-28 are the safest light aircraft. In the graph of fraction uninjured vs. model, we see these three aircrafts offer the best protection.

```
In [8]: dv.best_models(df)
```

/Users/jgoett/Library/CloudStorage/OneDrive-Personal/Documents/Fall2024/Flati
ron-Live/Phase1/Project/Phase1-Airplane-Zach-Jeff/scr/data_visualizations.py:
58: FutureWarning: Calling float on a single element Series is deprecated and
will raise a TypeError in the future. Use float(ser.iloc[0]) instead
  xy = pd.DataFrame([(x,float(ave_frac.loc[x])) for x in top_5]).sort_values(
by=1)

Destroyed per Make_Model in Ininjured
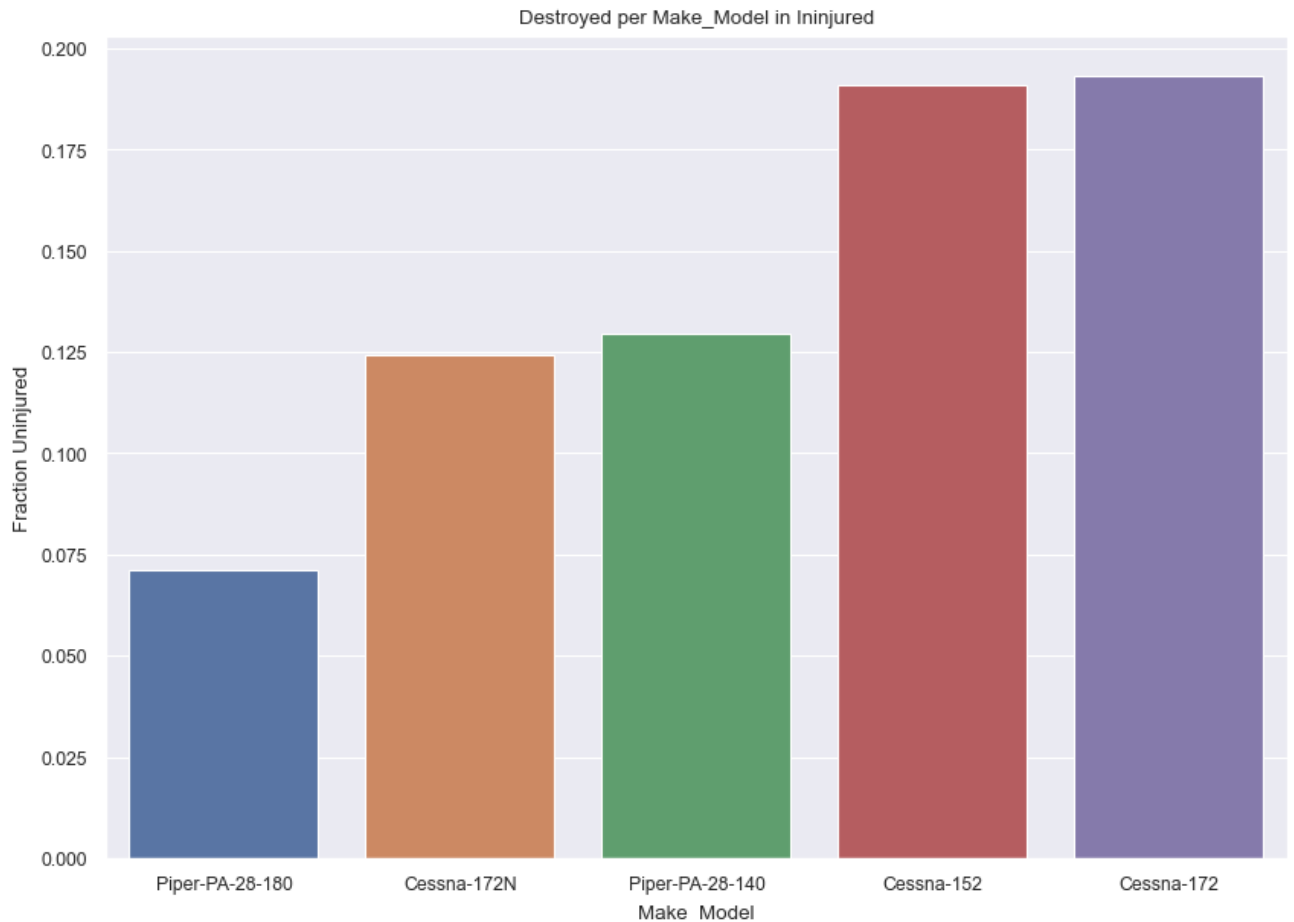
Airplane Safety during Phases of Flight

To help our students become safer pilots, we consider which phases of flight have historically lead to the most accident fatalities in light weight aircraft. Below, we consider the fraction of passengers killed during each phase of flight. We find the maneuvering, climbing, and cruising are the most dangerous phases.

```
In [9]: df["Broad.phase.of.flight"].value_counts().drop(["Unknown","Other"]).index
```

```
Out[9]: Index(['Landing', 'Takeoff', 'Cruise', 'Approach', 'Maneuvering', 'Climb',
               'Taxi', 'Descent', 'Go-around', 'Standing'],
              dtype='object', name='Broad.phase.of.flight')
```

```
In [10]: dv.performance_phase(df, "Fraction_fatal")
```

Fraction_fatal vs. Phase of Flight
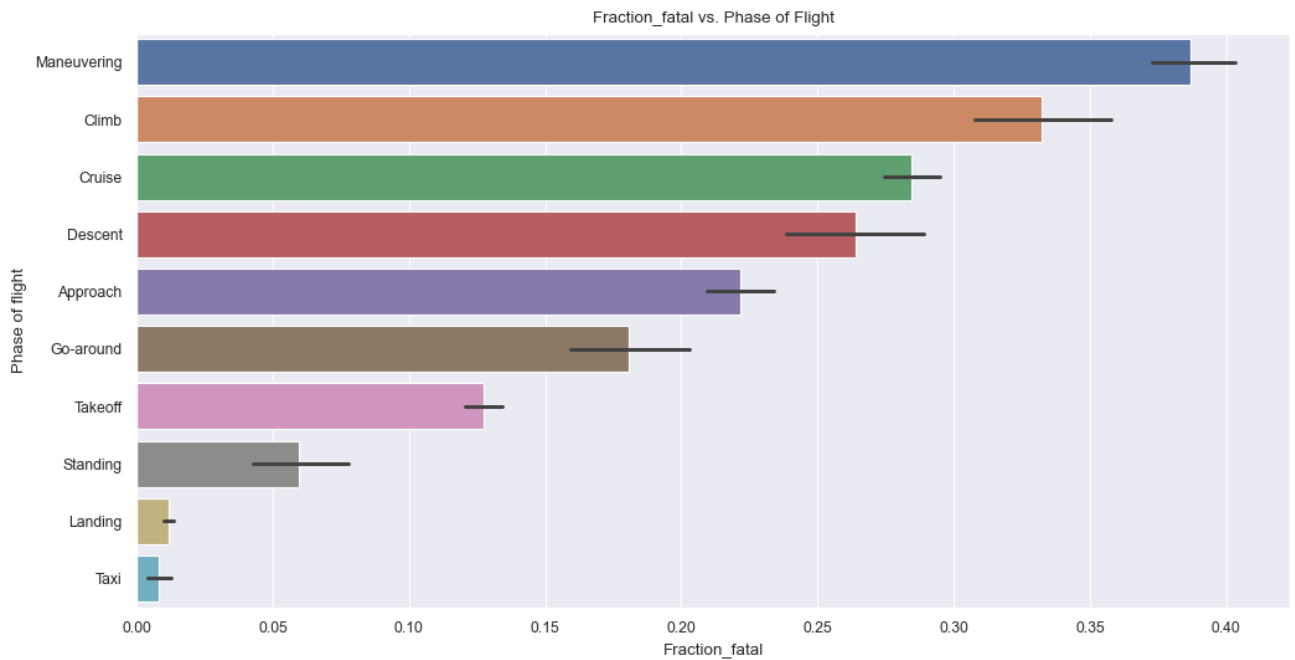
We discover similar results when considering an "inverse statistic", the fraction of passengers uninjured during accidents during each phase.

```
In [11]: dv.performance_phase(df, "Fraction_uninjured")
```


Fraction_uninjured vs. Phase of Flight
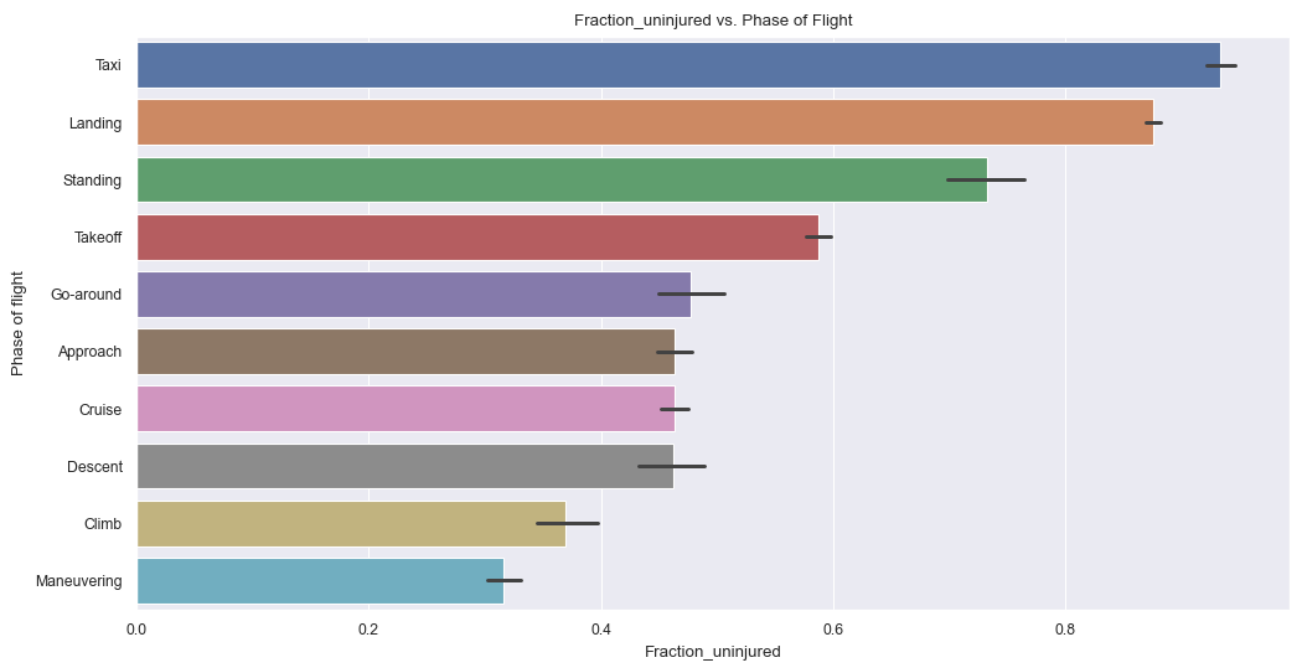
Calculating Heatmap

For fun, we break down how different aircraft make perform safety wise during different flight phases. Our safeest manufacturers, Cessna and Bell, both encounter their most dangerous phases during maneuvering.

```
In [12]: phases_flight = df["Broad.phase.of.flight"].value_counts().index[:-2]
         df_phases = df[df["Broad.phase.of.flight"].map(lambda x: x in phases_flight)
```

```
In [13]: fig, ax = plt.subplots(figsize=(10,6))


         table = pd.pivot_table(df_phases, index=['Make'],columns=['Broad.phase.of.fl
         sns.heatmap(table, cmap="Blues", cbar_kws={"label":"Fraction Uninjured"}, ax
         ax.set_title('Heatmap', fontsize=16, fontweight = 'bold')
         ax.tick_params(labelsize=12)
         ax.set_xlabel('Make', fontweight = 'bold', fontsize=14)
         ax.set_ylabel('Phase of flight', fontweight = 'bold', fontsize=14)

         fig.savefig("Images/Uninjured_Make_Phases", bbox_inches = 'tight')
```
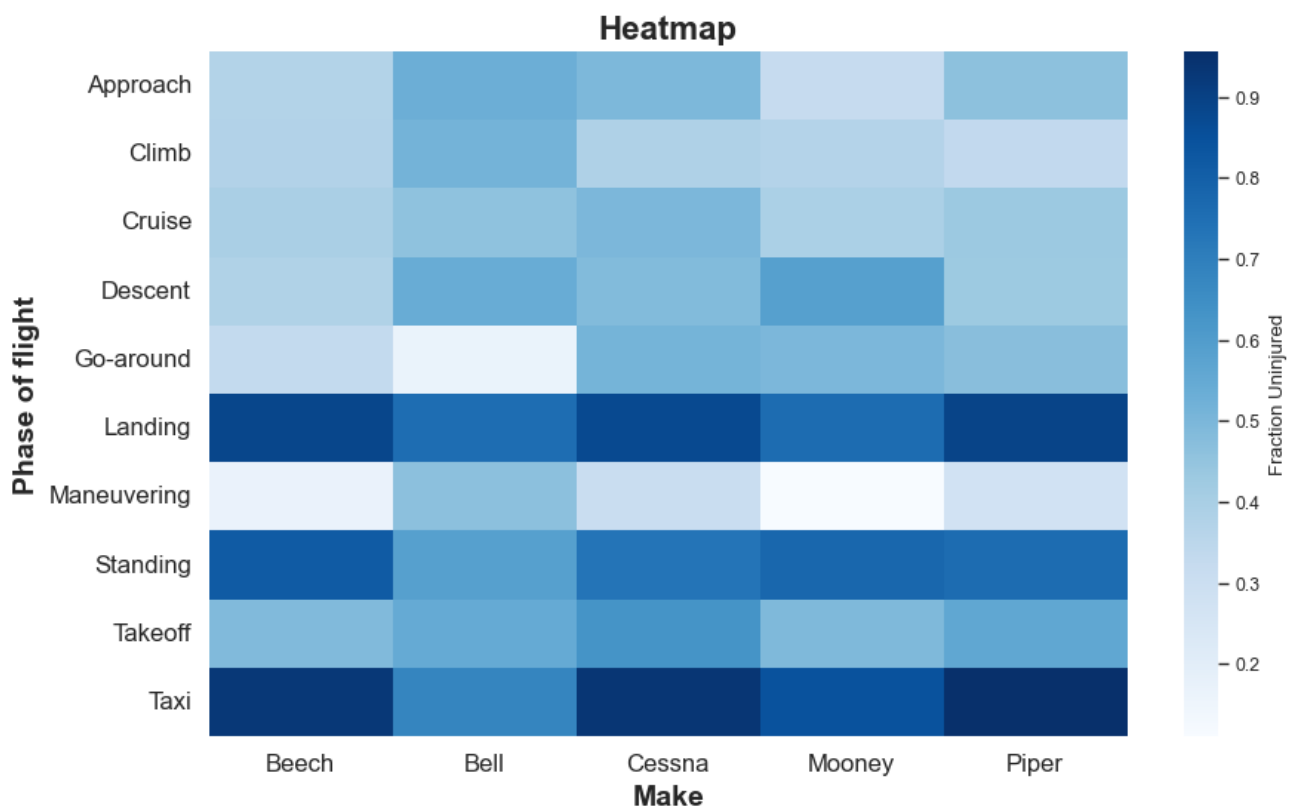

Airplane Safety during Different Months of the Year

Finally, to support our students, we investigate which months of the year are correlated to the most deadly accidents. In the graphs of month vs. fraction fatal and fraction uninjured below, we find that the winter months are correlated to higher fatality rates, and the summer months are correlated to the best survival rates.

In [14]:
```python
fig, ax = plt.subplots(1,2,figsize=[14,5])

sns.barplot(data=df, x='Month', y = 'Fraction_fatal', ax=ax[0])
sns.barplot(data=df, x='Month', y = 'Fraction_uninjured', ax=ax[1])

fig.savefig("Images/Survival_month")
```
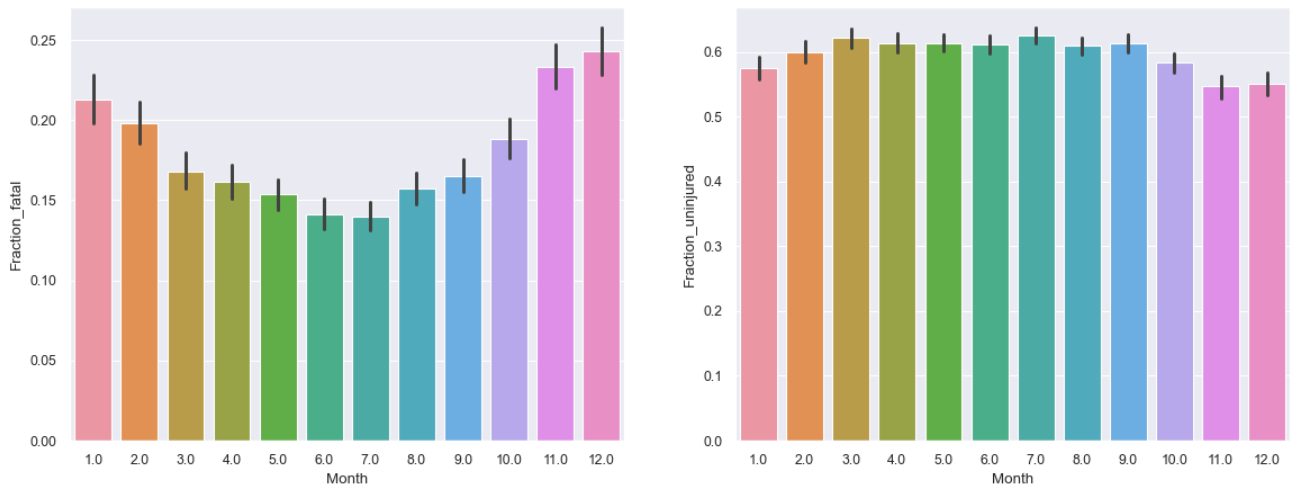
Conclusions

- Consider **Cessna** or **Piper** makes (Cessna 172, 158, or Piper PA-28)
- Focus on safety during **climb**, **cruize**, and **maneuvering** phases of flight
- Focus on safety during **winter** flying.

Next Steps

- Investigate the geographical locations of accidents to gain insight.
- Investigate weather conditions.