In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Show plots inline and set style
%matplotlib inline
sns.set(style="whitegrid")

# Load data
df = pd.read_csv("airline_accidents.csv", low_memory=False)

# Clean column names
df.columns = (
    df.columns
    .str.strip()
    .str.lower()
    .str.replace(" ", "_")
    .str.replace(r'[^\w\s]', '', regex=True)  # remove special chars
)

# Handle missing values
df = df.dropna(subset=['make', 'model'])   # drop rows missing make/model
df = df.fillna(0)                          # fill other NaNs with 0

# Quick summary
print("Dataset shape:", df.shape)
print("\nFirst 5 rows:")
display(df.head())
print("\nColumn info:")
df.info()
```

```
Dataset shape: (150959, 31)

First 5 rows:
```

	event_id	investigation_type	accident_number	event_date	location	country
0	20080125X00106	Accident	SEA08CA056	12/31/2007	Santa Ana, CA	United States
1	20080206X00141	Accident	CHI08WA075	12/31/2007	Guernsey, United Kingdom	United Kingdom
2	20080129X00122	Accident	CHI08CA057	12/30/2007	Alexandria, MN	United States
3	20080114X00045	Accident	LAX08FA043	12/30/2007	Paso Robles, CA	United States
4	20080109X00032	Accident	NYC08FA071	12/30/2007	Cherokee, AL	United States

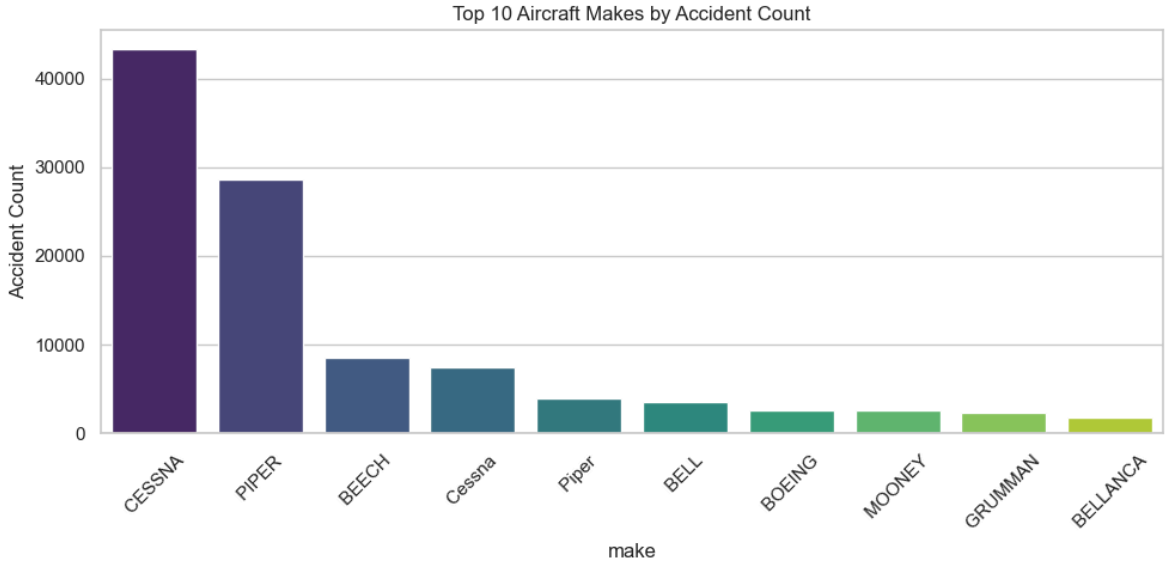
5 rows × 31 columns

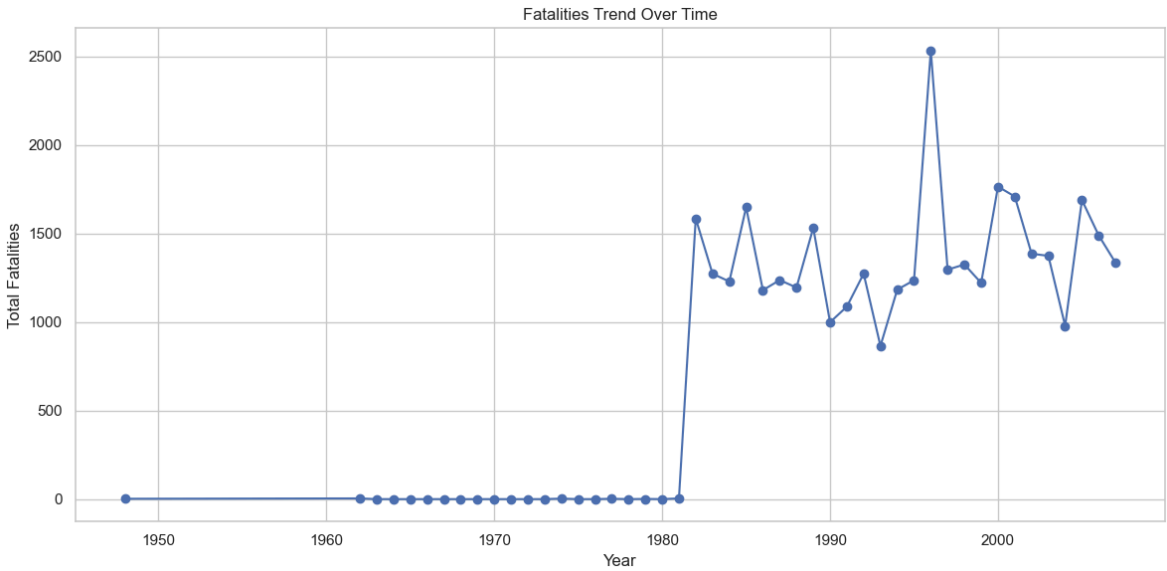
```
Column info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150959 entries, 0 to 150958
Data columns (total 31 columns):
 #   Column                  Non-Null Count   Dtype 
---  ------                  --------------   ----- 
 0   event_id                150959 non-null  object
 1   investigation_type      150959 non-null  object
 2   accident_number         150959 non-null  object
 3   event_date              150959 non-null  object
 4   location                150959 non-null  object
 5   country                 150959 non-null  object
 6   latitude                150959 non-null  object
 7   longitude               150959 non-null  object
 8   airport_code            150959 non-null  object
 9   airport_name            150959 non-null  object
 10  injury_severity         150959 non-null  object
 11  aircraft_damage         150959 non-null  object
 12  aircraft_category       150959 non-null  object
 13  registration_number     150959 non-null  object
 14  make                    150959 non-null  object
 15  model                   150959 non-null  object
 16  amateur_built           150959 non-null  object
 17  number_of_engines       150959 non-null  object
 18  engine_type             150959 non-null  object
 19  far_description         150959 non-null  object
 20  schedule                150959 non-null  object
 21  purpose_of_flight       150959 non-null  object
 22  air_carrier             150959 non-null  object
 23  total_fatal_injuries    150959 non-null  object
 24  total_serious_injuries  150959 non-null  object
 25  total_minor_injuries    150959 non-null  object
 26  total_uninjured         150959 non-null  object
 27  weather_condition       150959 non-null  object
 28  broad_phase_of_flight   150959 non-null  object
 29  report_publication_date 150959 non-null  object
 30  unnamed_30              150959 non-null  object
dtypes: object(31)
memory usage: 35.7+ MB
```

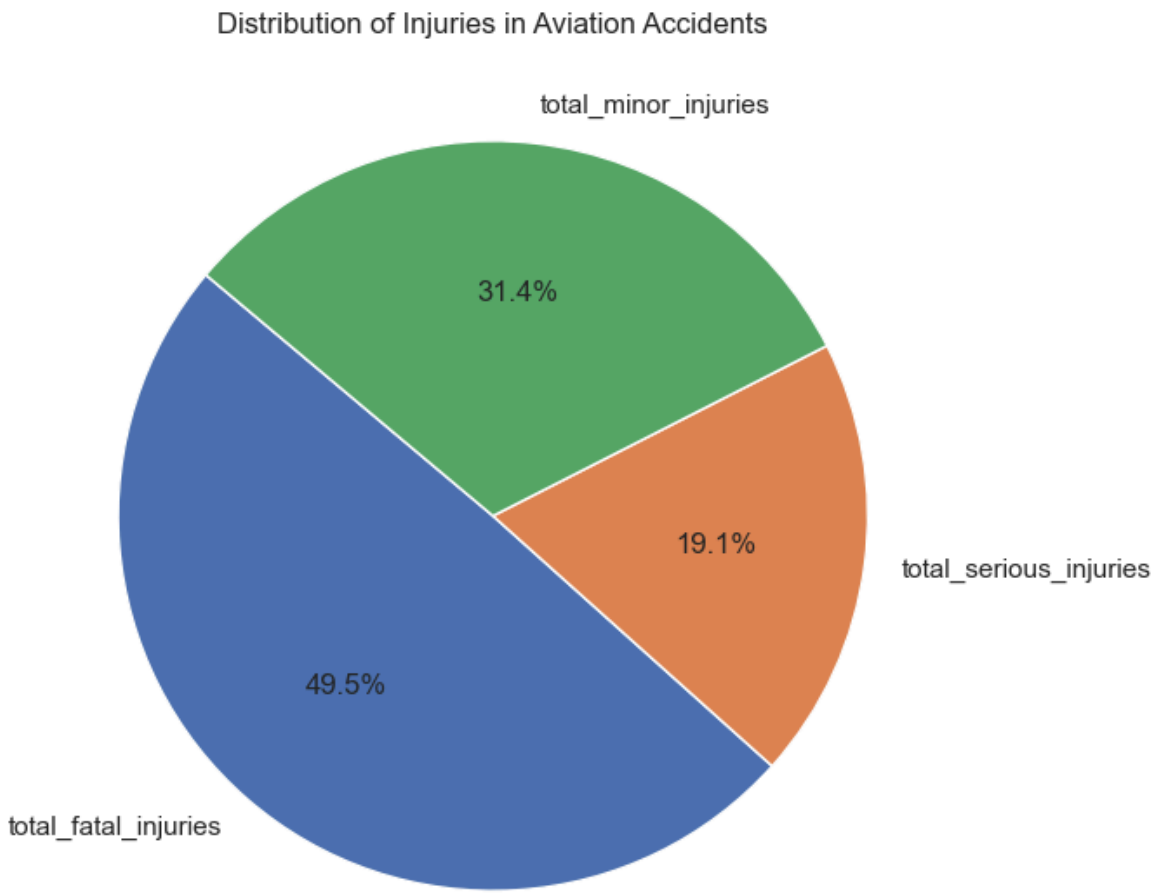
In [2]:

```python
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# --- 1. Top 10 Aircraft Makes ---
top_makes = df['make'].value_counts().head(10)
plt.figure(figsize=(10,5))
sns.barplot(x=top_makes.index, y=top_makes.values, palette="viridis")
plt.title("Top 10 Aircraft Makes by Accident Count")
plt.ylabel("Accident Count")
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig("top10_makes.png")
plt.show()

# --- 2. Fatalities Trend Over Time ---
df['event_date'] = pd.to_datetime(df['event_date'], errors='coerce')
df['total_fatal_injuries'] = pd.to_numeric(df['total_fatal_injuries'], errors='c

fatalities_trend = df.groupby(df['event_date'].dt.year)['total_fatal_injuries'].
```

```
plt.figure(figsize=(12,6))
fatalities_trend.plot(kind='line', marker='o')
plt.title("Fatalities Trend Over Time")
plt.ylabel("Total Fatalities")
plt.xlabel("Year")
plt.grid(True)
plt.tight_layout()
plt.savefig("fatalities_trend.png")
plt.show()

# --- 3. Distribution of Injuries ---
injuries = df[['total_fatal_injuries','total_serious_injuries','total_minor_inju

plt.figure(figsize=(7,7))
plt.pie(injuries, labels=injuries.index, autopct='%1.1f%%', startangle=140)
plt.title("Distribution of Injuries in Aviation Accidents")
plt.savefig("injury_distribution.png")
plt.show()
```






Distribution of Injuries in Aviation Accidents
total_minor_injuries
31.4%
19.1%
total_serious_injuries
49.5%
total_fatal_injuries

In [ ]: