**SAS Studios**

FILENAME REFFILE '/folders/myfolders/sasuser.v94/airline\_delay.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=WORK.IMPORT1;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=WORK.IMPORT1; RUN;

Summary statistics

proc means data=WORK.IMPORT1 chartype mean std min max n vardef=df;

var \_arr\_delay \_carrier\_delay weather\_delay nas\_delay security\_delay

late\_aircraft\_delay;

class year;

run;

proc means data=WORK.IMPORT1 chartype mean std min max n vardef=df;

var \_arr\_delay \_carrier\_delay weather\_delay nas\_delay security\_delay

late\_aircraft\_delay;

class carrier;

Run;

proc means data=WORK.IMPORT1 chartype mean std min max n vardef=df;

var arr\_flights arr\_del15 arr\_cancelled arr\_diverted \_arr\_delay;

class year;

Run

proc means data=WORK.IMPORT chartype mean std min max n sum vardef=df;

var arr\_flights arr\_del15 arr\_cancelled arr\_diverted;

class year;

Run;

**PYTHON JUPYTER**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

# load Alaska Airlines data set

filename = 'airline\_delays.csv'

df = pd.read\_csv(filename)

df.head()

# Total number of flights per category per year

df.groupby('year').sum()[['arr\_flights', 'arr\_del15', 'arr\_cancelled', 'arr\_diverted']]

df.groupby('year').sum()[['arr\_flights', 'arr\_del15', 'arr\_cancelled', 'arr\_diverted']].plot(kind='bar')

plt.show

# Overall total number of flights per category

df.sum()[['arr\_flights', 'arr\_del15', 'arr\_cancelled', 'arr\_diverted']]

#pie chart of results

labels = ['arr\_flights', 'arr\_del15', 'arr\_cancelled', 'arr\_diverted']

sizes = [332110, 60455, 6938, 677]

plt.pie(sizes, labels=labels, autopct='%1.1f%%')

plt.title("Number of flights per category")

plt.show

# Overall average number of airline delays per category

df.mean() [[' carrier\_delay', 'weather\_delay', 'nas\_delay', 'security\_delay', 'late\_aircraft\_delay']]

labels = ['nas\_delay','late\_aircraft\_delay',' carrier\_delay','weather\_delay','security\_delay']

sizes = [1926.17, 1555.17, 1261.98, 123.89, 9.20]

plt.pie(sizes, labels=labels, autopct='%1.1f%%')

plt.title("Type of delays")

plt.show

# average number of flights per delay cause per year

df.groupby('year').mean() [[' carrier\_delay','weather\_delay','nas\_delay','security\_delay','late\_aircraft\_delay']]

df.groupby('year').mean() [[' carrier\_delay','weather\_delay','nas\_delay','security\_delay','late\_aircraft\_delay']].plot(kind='bar')

plt.show