spark

July 1, 2021

```
[ ]: t  = spark.read.csv("gs://dataproc-staging-europe-west4-1041851436623-ceaakthz/
     ↪y",header=True, inferSchema=True)
```

```
[ ]:
```

```
[ ]:
```

```
[ ]: import pandas
     from collections import Counter
     wee=[]
     for i in range(22):
         ty= t.filter(t.Year == 1987+i)
         ty1=ty.groupby('UniqueCarrier').count()
         a=ty1.sort("count", ascending=False).take(10)

         we=[]
         for k in range(10):
             s=[a[k][0]]*a[k][1]
             we.extend(s)
         wee.append(we)
         letter_counts = Counter(we)
         df = pandas.DataFrame.from_dict(letter_counts, orient='index')
         df.plot(title=str(1987+i),kind='bar')
```

```
[287]: for i in range(22):

           letter_counts = Counter(wee[i])
           df = pandas.DataFrame.from_dict(letter_counts, orient='index')
           df.plot(title=str(1987+i),kind='bar')
```

1987

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2003

2004

2005

2006