

```

import time
from pyspark.sql.functions import sum

df = spark.read.csv("s3://thathsara-video-presentation-
assignment/input/DelayedFlights.csv", header=True,
inferSchema=True)
df.count()

df.createOrReplaceTempView("AirDelay")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").agg(
    {"CarrierDelay": "sum", "NASDelay": "sum", "WeatherDelay":
"sum", "LateAircraftDelay": "sum", "SecurityDelay": "sum"
}).show()
end_time = time.time()
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").sum("CarrierDelay").show()
end_time = time.time()
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").sum("NASDelay").show()
end_time = time.time()
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").sum("WeatherDelay").show()
end_time = time.time()
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").sum("LateAircraftDelay").show()
end_time = time.time()

```

```
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")

start_time = time.time()
process_df = df.filter((df.Year >= 2003) & (df.Year <=
2010)).groupBy("Year").sum("SecurityDelay").show()
end_time = time.time()
elapsed_time = end_time - start_time
print(f"Elapsed time: {elapsed_time:.2f} seconds")
```