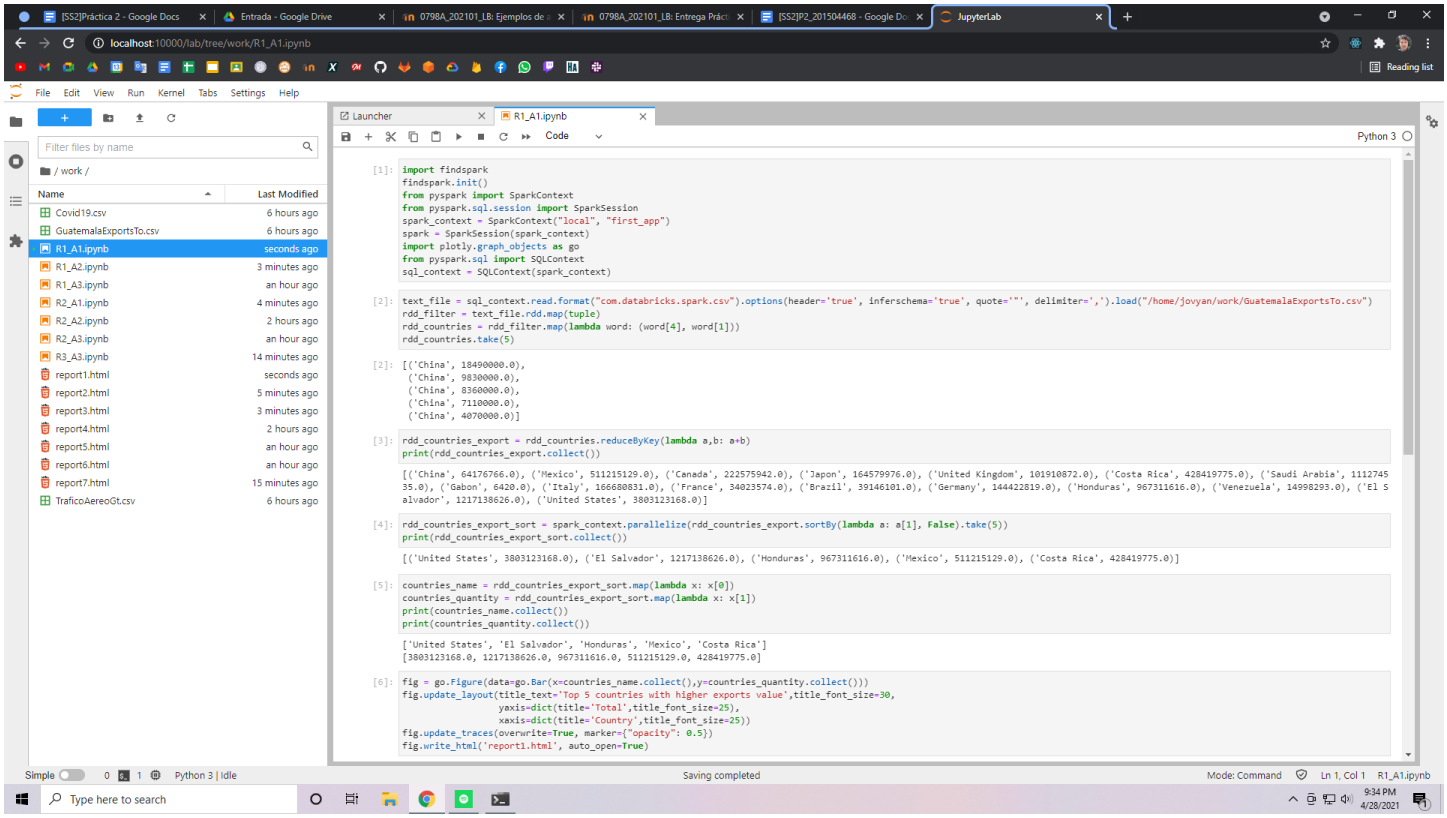
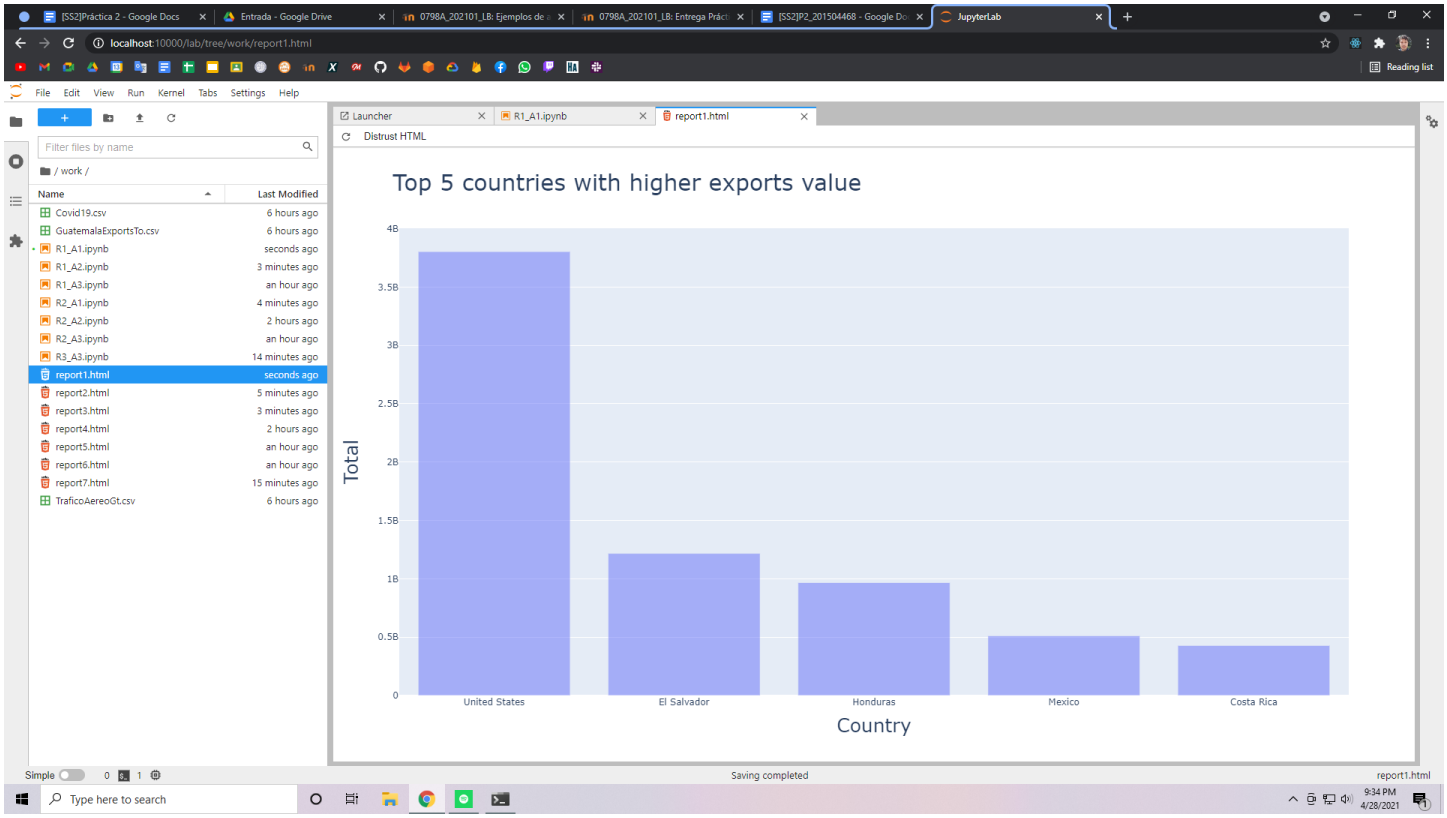
Práctica 2

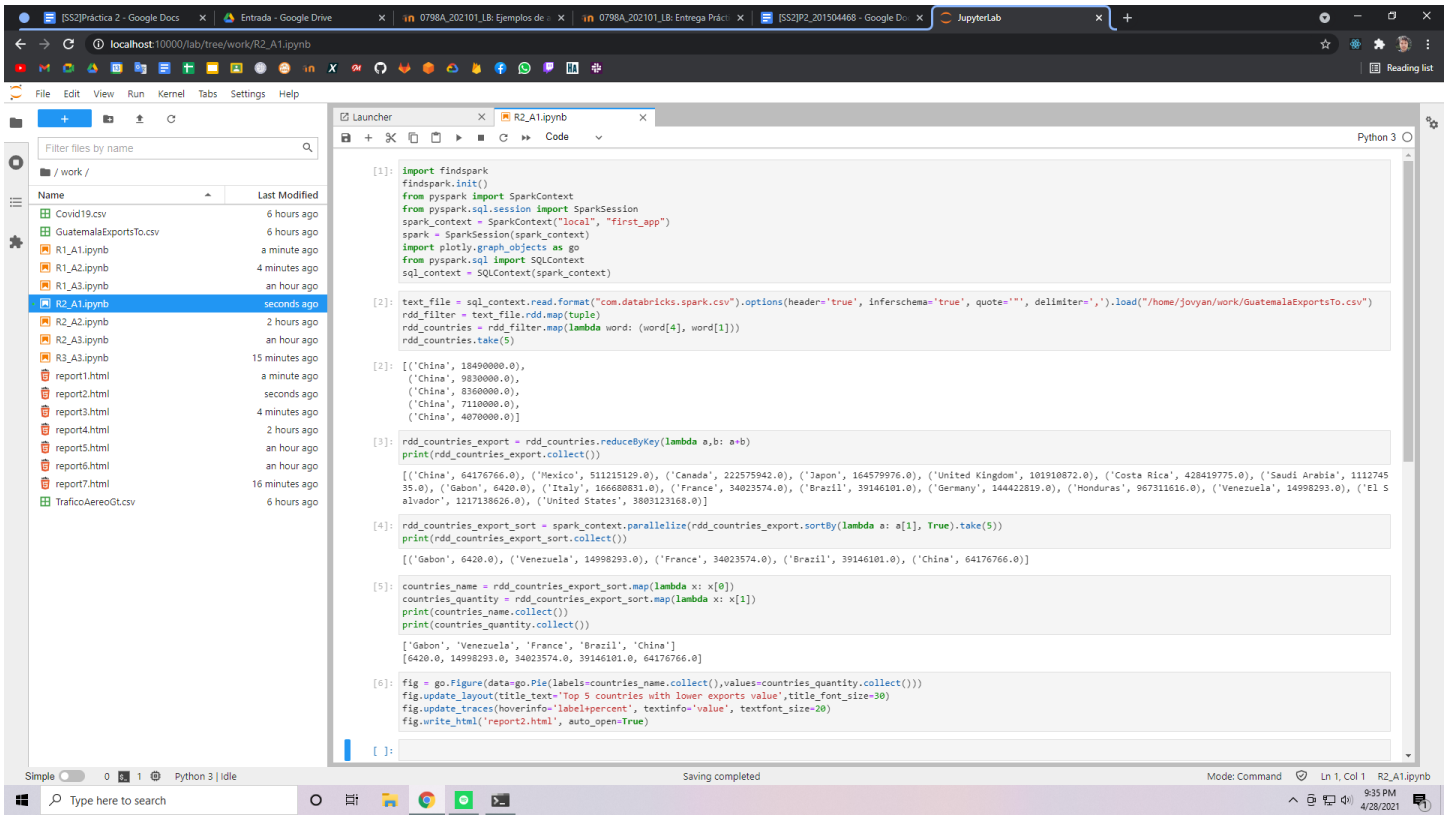
Archivo 1 - “GuatemalaExportsTo.csv”

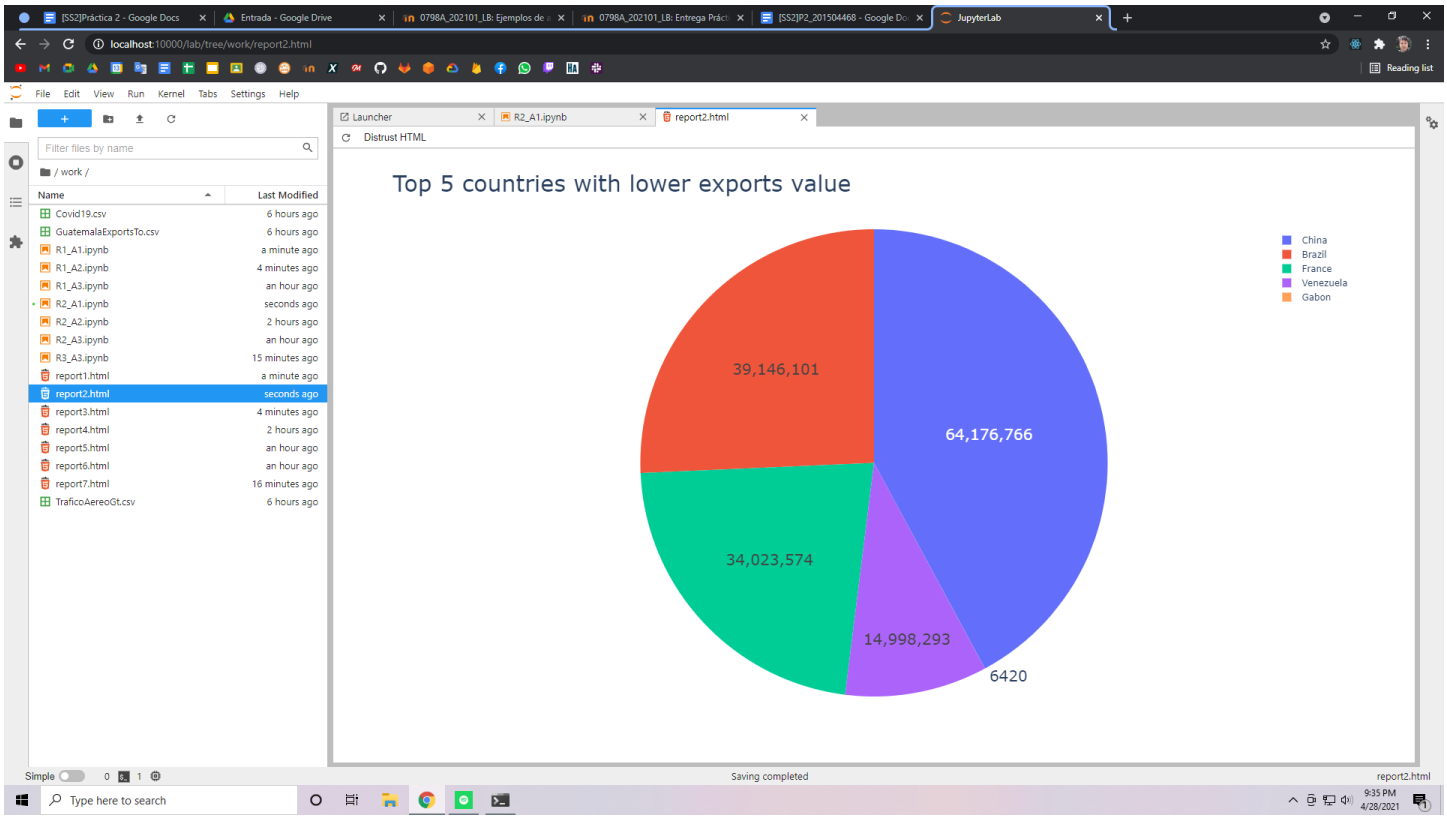
1. Gráfica de barras que muestre el país con el mayor valor por exportaciones.





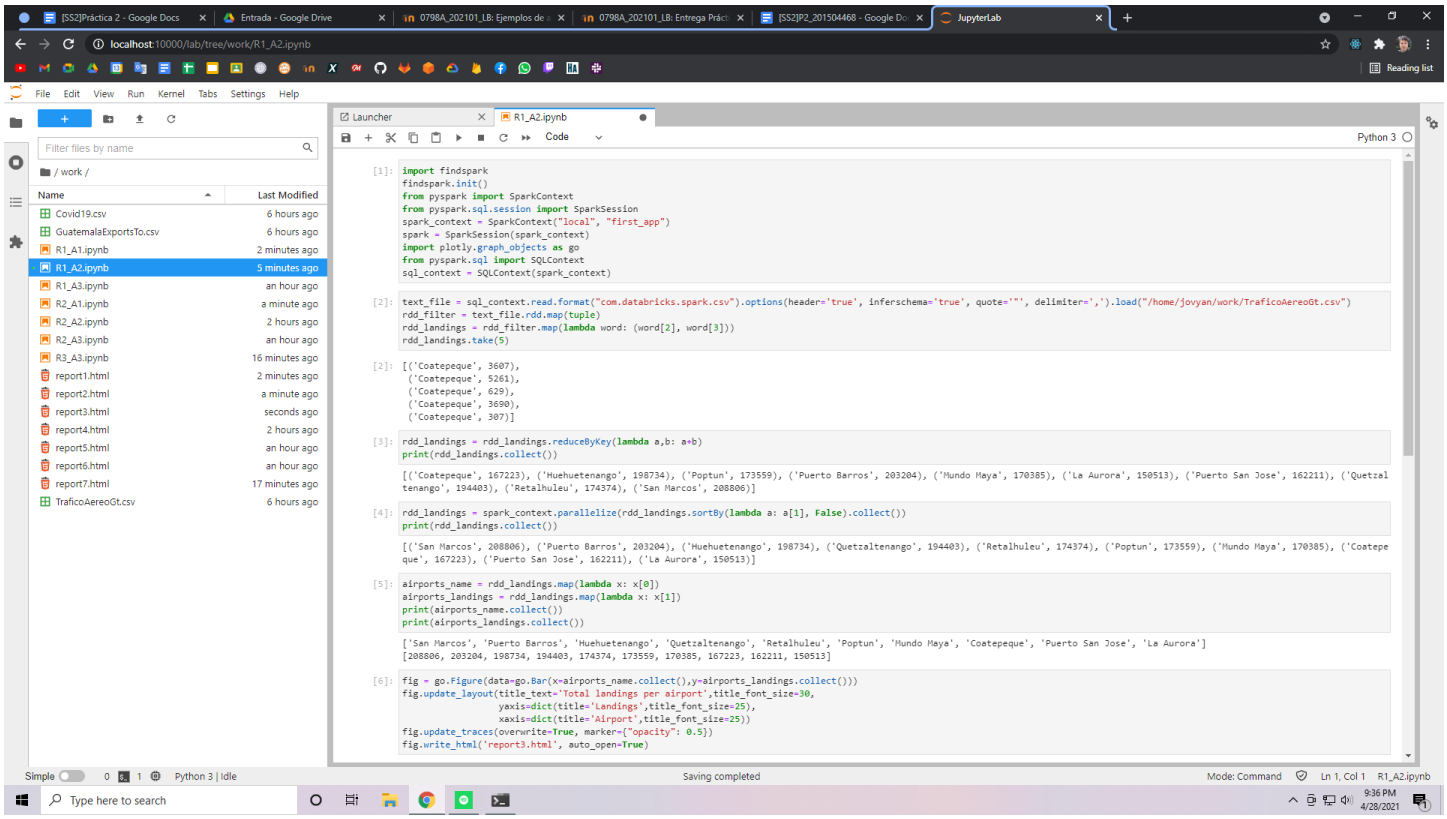
2. Gráfica de pie de los 5 países con menos valor por exportaciones.

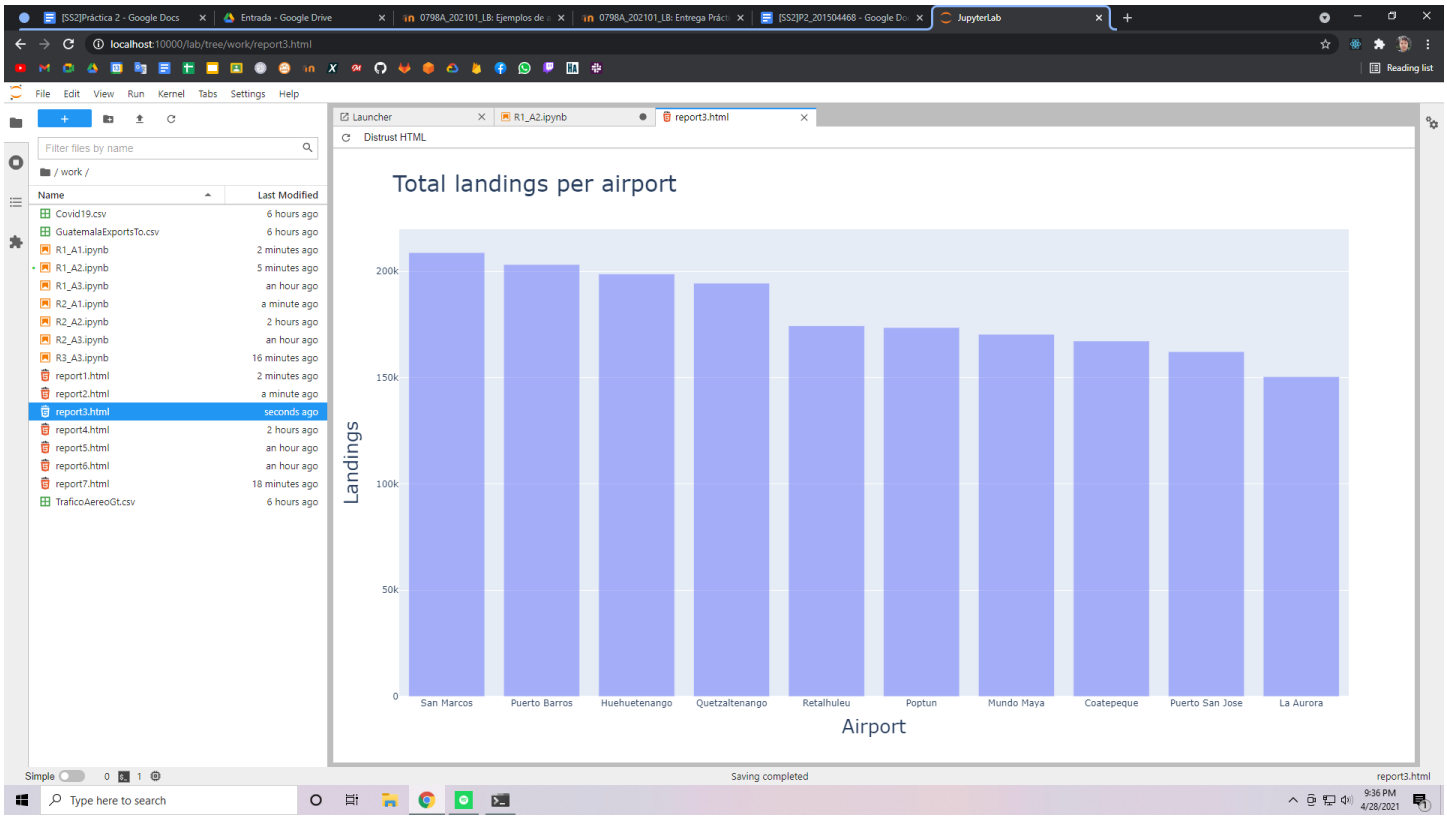




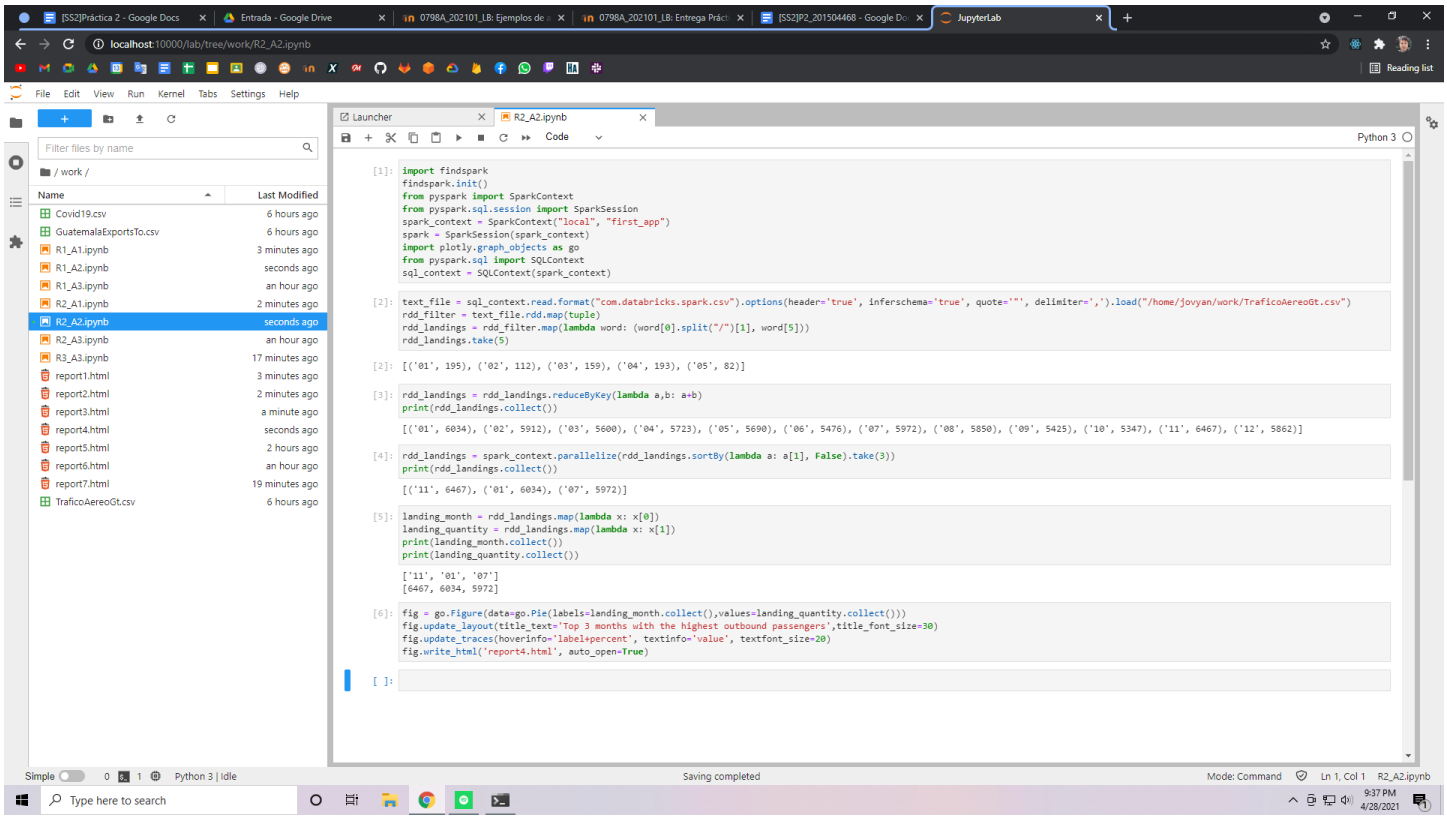
Archivo 2 - “TraficoAereoGt.csv”

1. Gráfica de barra con el total de aterrizajes por Aeropuerto.





2. Gráfica de pie con los 3 meses con mayor número de pasajeros de salida.



```
import findspark
findspark.init()
from pyspark import SparkContext
from pyspark.sql.session import SparkSession
spark_context = SparkContext("local", "first_app")
spark = SparkSession(spark_context)
import plotly.graph_objects as go
from pyspark.sql import SQLContext
sql_context = SQLContext(spark_context)
```

```
text_file = sql_context.read.format("com.databricks.spark.csv").options(header='true', inferschema='true', quote='"', delimiter=',').load("/home/jovyan/work/TraficoAereoGt.csv")
rdd_filter = text_file.rdd.map(tuple)
rdd_landings = rdd_filter.map(lambda word: (word[0].split("/")[1], word[5]))
rdd_landings.take(5)
```

```
[('01', 195), ('02', 112), ('03', 159), ('04', 193), ('05', 82)]
```

```
rdd_landings = rdd_landings.reduceByKey(lambda a,b: a+b)
print(rdd_landings.collect())
```

```
[('01', 6034), ('02', 5912), ('03', 5600), ('04', 5723), ('05', 5690), ('06', 5476), ('07', 5972), ('08', 5850), ('09', 5425), ('10', 5347), ('11', 6467), ('12', 5862)]
```

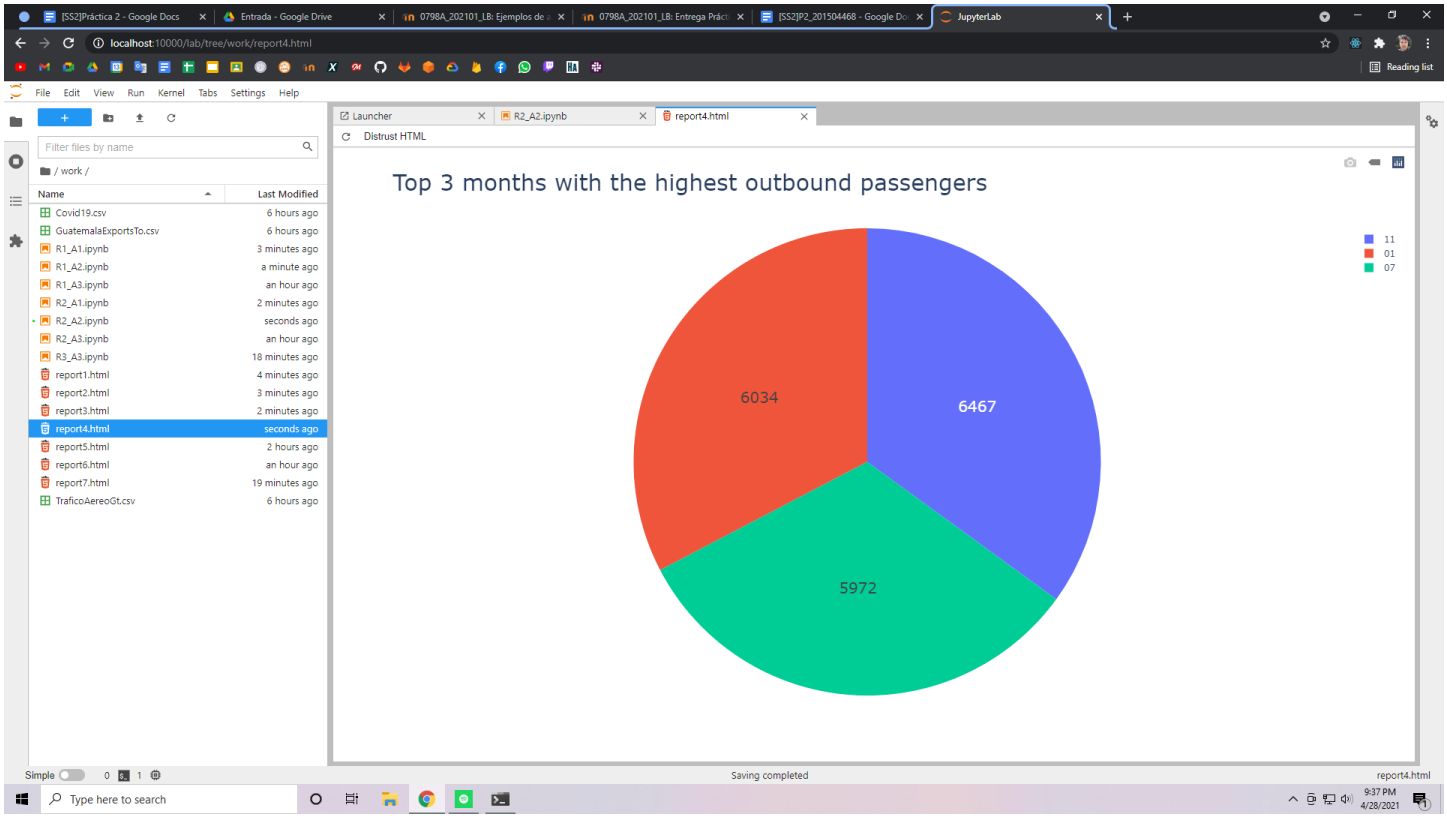
```
rdd_landings = spark_context.parallelize(rdd_landings.sortBy(lambda a: a[1], False).take(3))
print(rdd_landings.collect())
```

```
[('11', 6467), ('01', 6034), ('07', 5972)]
```

```
landing_month = rdd_landings.map(lambda x: x[0])
landing_quantity = rdd_landings.map(lambda x: x[1])
print(landing_month.collect())
print(landing_quantity.collect())
```

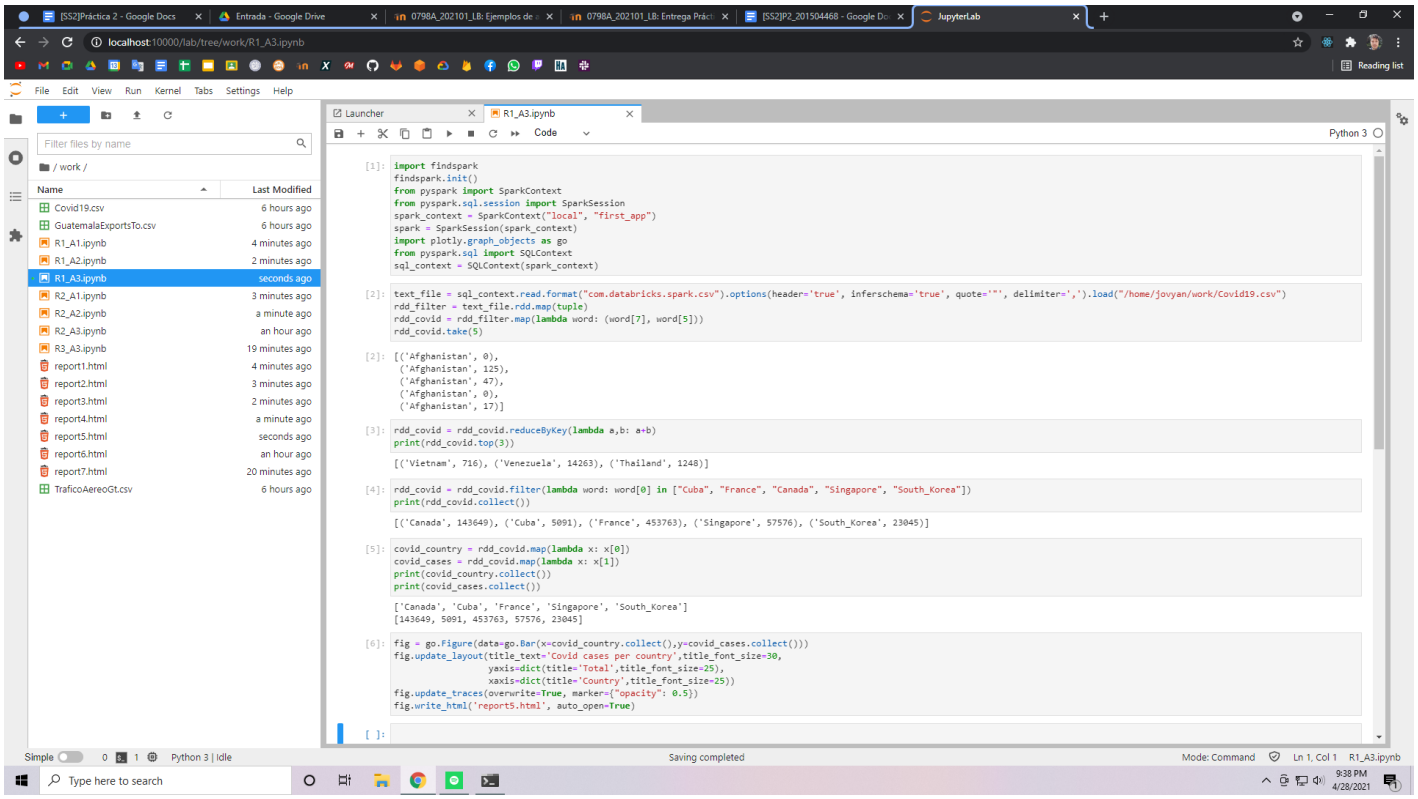
```
['11', '01', '07']
[6467, 6034, 5972]
```

```
fig = go.Figure(data=go.Pie(labels=landing_month.collect(),values=landing_quantity.collect()))
fig.update_layout(title_text='Top 3 months with the highest outbound passengers',title_font_size=30)
fig.update_traces(hoverinfo='label+percent', textinfo='value', textfont_size=20)
fig.write_html('report4.html', auto_open=True)
```



Archivo 3 - “Covid19”

1. Gráfica de barras con el número total de casos de Covid-19 en
    a. Cuba
    b. France
    c. Canada
    d. Singapore
    e. South_Korea



```
import findspark
findspark.init()
from pyspark import SparkContext
from pyspark.sql.session import SparkSession
spark_context = SparkContext("local", "first_app")
spark = SparkSession(spark_context)
import plotly.graph_objects as go
from pyspark.sql import SQLContext
sql_context = SQLContext(spark_context)
```

```
text_file = sql_context.read.format("com.databricks.spark.csv").options(header='true', inferschema='true', quote='"', delimiter=',').load("/home/jovyan/work/Covid19.csv")
rdd_filter = text_file.rdd.map(tuple)
rdd_covid = rdd_filter.map(lambda word: (word[7], word[5]))
rdd_covid.take(5)
```

```
[('Afghanistan', 0),
 ('Afghanistan', 125),
 ('Afghanistan', 47),
 ('Afghanistan', 0),
 ('Afghanistan', 17)]
```

```
rdd_covid = rdd_covid.reduceByKey(lambda a,b: a+b)
print(rdd_covid.top(3))
```

```
[('Vietnam', 716), ('Venezuela', 14263), ('Thailand', 1248)]
```

```
rdd_covid = rdd_covid.filter(lambda word: word[0] in ["Cuba", "France", "Canada", "Singapore", "South_Korea"])
print(rdd_covid.collect())
```

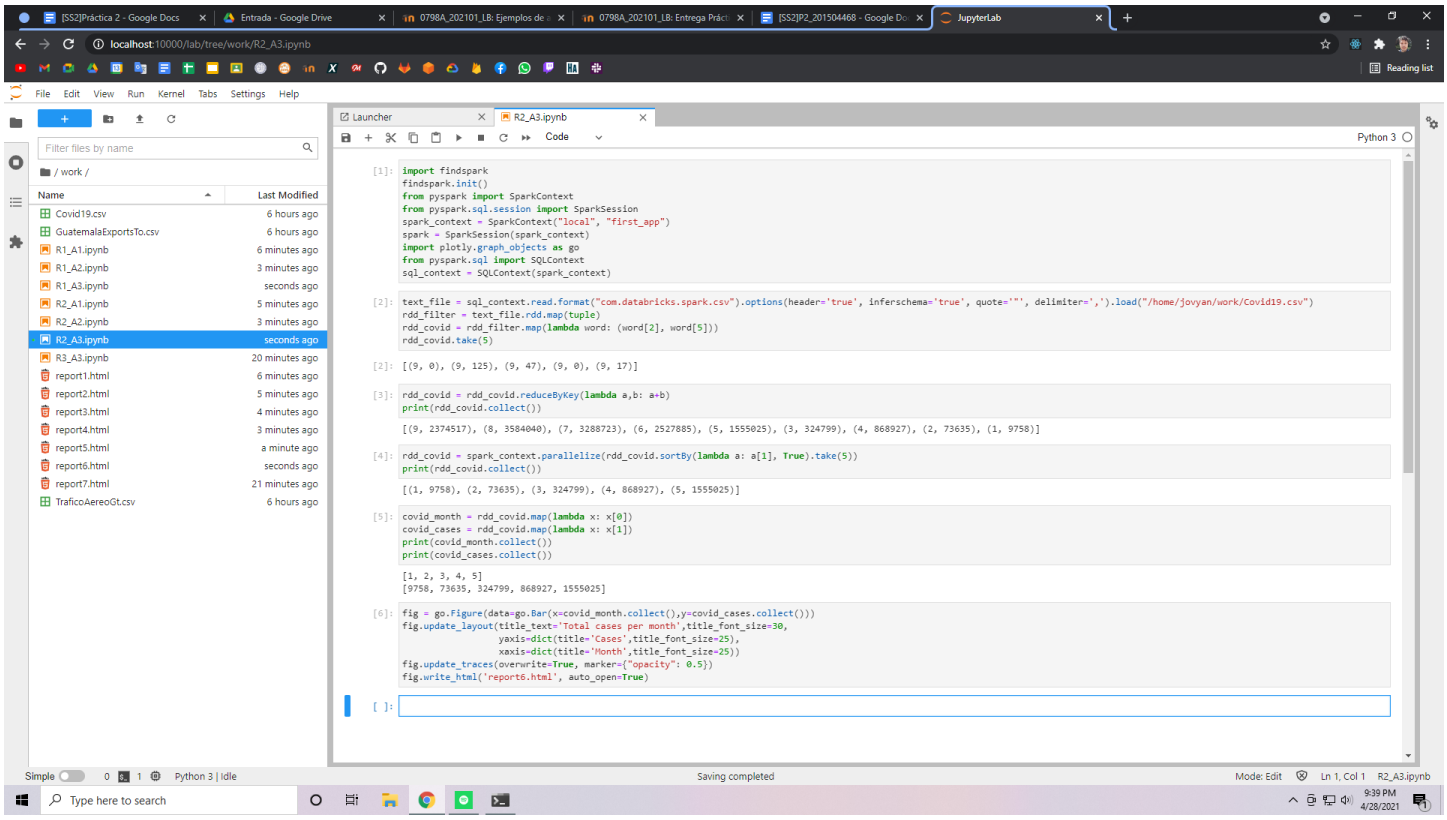
```
[('Canada', 143649), ('Cuba', 5091), ('France', 453763), ('Singapore', 57576), ('South_Korea', 23045)]
```

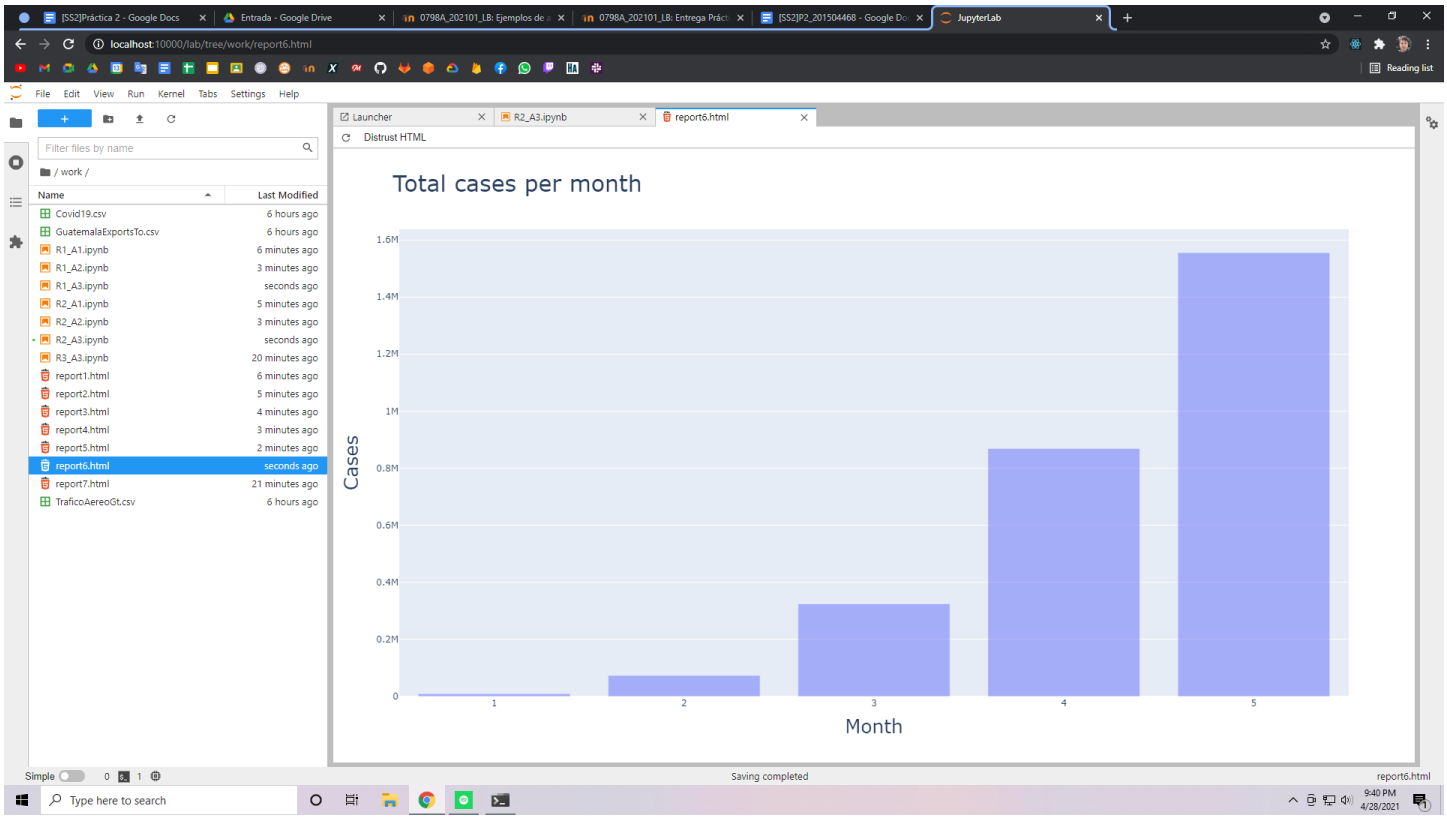
```
covid_country = rdd_covid.map(lambda x: x[0])
covid_cases = rdd_covid.map(lambda x: x[1])
print(covid_country.collect())
print(covid_cases.collect())
```

```
['Canada', 'Cuba', 'France', 'Singapore', 'South_Korea']
[143649, 5091, 453763, 57576, 23045]
```

```
fig = go.Figure(data=go.Bar(x=covid_country.collect(),y=covid_cases.collect()))
fig.update_layout(title_text='Covid cases per country',title_font_size=30,
                  yaxis=dict(title='Total',title_font_size=25),
                  xaxis=dict(title='Country',title_font_size=25))
fig.update_traces(overwrite=True, marker={"opacity": 0.5})
fig.write_html('report5.html', auto_open=True)
```

2. Gráfica de pie con los 5 meses con menos casos de Covid-19.





3. Gráfica a su elección del total de casos y muertes en Guatemala en el mes de Agosto.