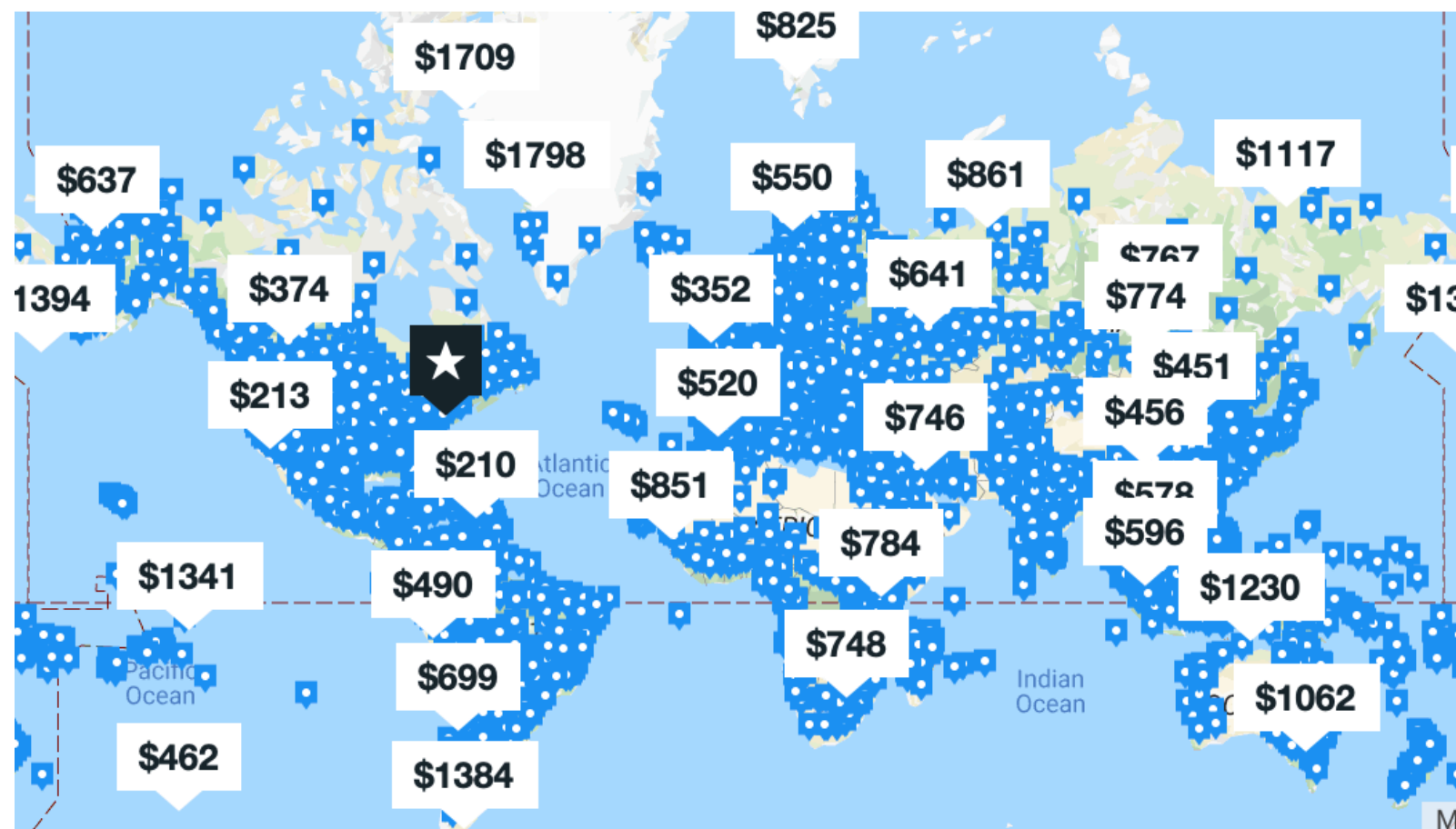# Factors responsible for Airline Pricing

06 August, 2019

Why?

Airline Industry is one of the biggest contributors to US economy.

Identifying the causes for flight prices is still pretty elusive.

Understanding how airlines structure their prices can allow users to strategize their purchase.

Flights   Hotels   Cars   Packages   Cruises   Deals   More ⌄

Round-trip ⌄   1 Adult ⌄   Economy ⌄   0 Bags ⌄

New York (NYC) ✕   ⇄   Los Angeles (LAX) ✕   📅 Fri 9/6 ‹ › | 📅 Tue 9/10 ‹ ›   🔍

OUR ADVICE 📈
Buy now

Prices are unlikely to decrease within 7 days ⓘ

Track prices ⚪ OFF

1824 of **2004 flights**

Fee Assistant ⓘ ⌃

🧳 Carry-on bag   ⊟ 0 ⊞
🧳 Checked bag   ⊟ 0 ⊞ **TO**
                        FROM

Stops

☑ Nonstop $237
☑ 1 stop $181
☑ 2+ stops $229

Times

| Take-off | Landing |

Take-off from NYC
Fri 5:00 AM – 11:00 PM

Cheapest	Best ⓘ	Quickest	Other Sort
$181 • 6h 47m	$181 • 6h 47m	$260 • 5h 20m	

Delta® Flights from New York to Los Angeles

DELTA   Enjoy Free In-Flight Entertainment on Delta Flights. Terms & Bag Fees Apply.
$237   nonstop
delta.com | Sponsored

$237
Delta

[View Deal]

Cheapest   🌙 Rating: 6   🧳✕ 🧳0

☐ 🌐 7:00 am — 9:45 am   nonstop   5h 45m $181
     United Airlines   EWR - LAX   **Hacker Fare**
                                                 Basic Economy
☐ spirit 12:17 pm — 11:07 pm 1 stop   7h 50m
     Spirit Airlines   MSY   LAX - EWR   [View Deal]

Best   ⭕ Rating: 10   🧳✕ 🧳0

☐ 🌐 7:00 am — 9:45 am   nonstop   5h 45m $237
     United Airlines   EWR - LAX   United Airlines
                                                 Basic Economy
☐ 🌐 9:15 pm — 5:15 am⁺¹   nonstop   5h 00m
     United Airlines   LAX - EWR   [View Deal]
                                              Economy $307

How?

https://www.kayak.com/flights/NYC-LAX/2019-09-06/2019-09-10?sort=bestflight_a
https://www.kayak.com/flights/NYC-BOS/2019-09-06/2019-09-10?sort=bestflight_a
https://www.kayak.com/flights/NYC-PER/2019-09-06/2019-09-10?sort=bestflight_a
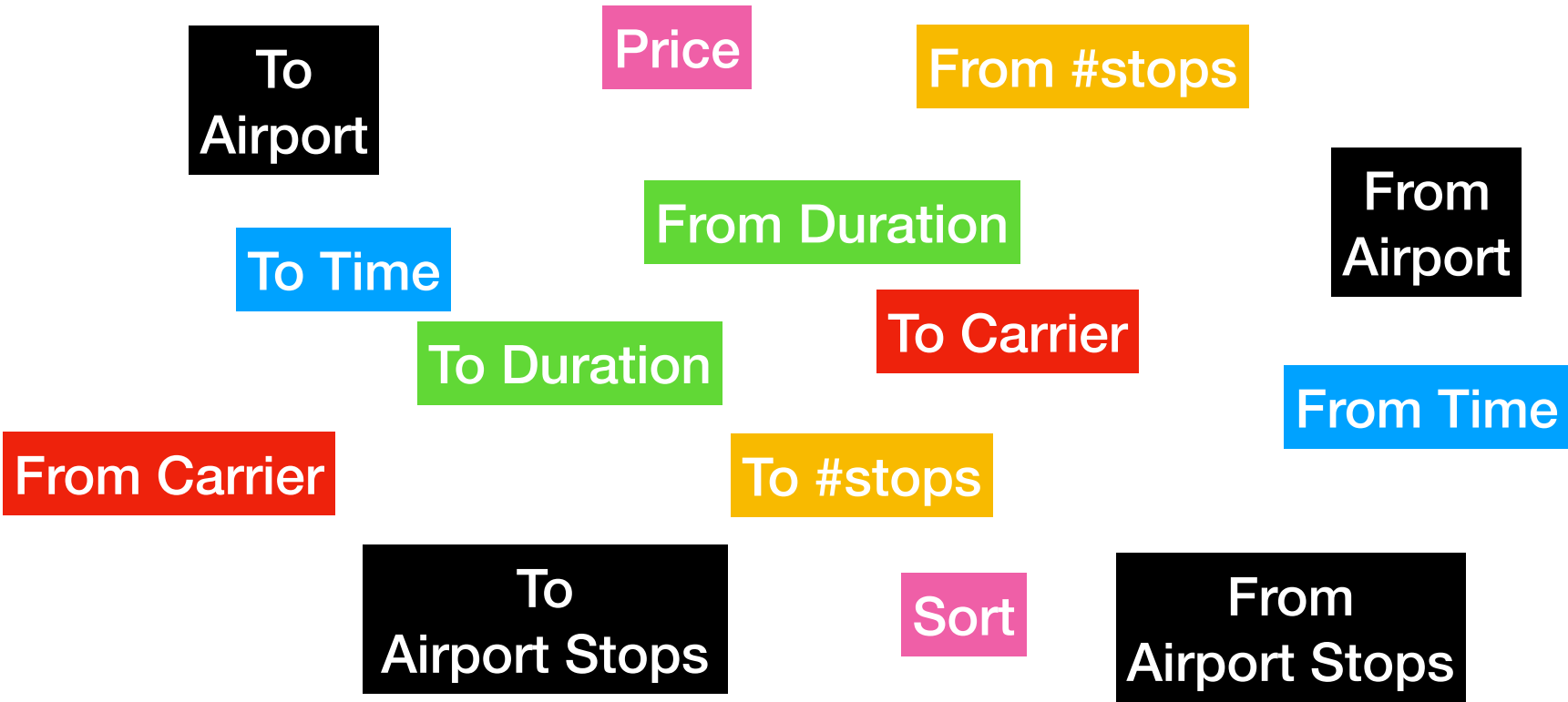
WebScrape

```python
def makeFlightURL(self,source,dest,start_date,end_date):
    return f"https://www.kayak.com/flights/{source}-{dest}/{start_date}/{end_date}?sort=bestflight_a"
```

```python
def scrape_sortings(self):
    """ returns scraped data in 3-tuple of differents sorting methods on flight page
        in order (best, cheapest, quickest)"""
    return [self.scrape_data(),
            self.scrape_data("cheapest"),
            self.scrape_data("quickest")]
```

```python
def scrape_flight(self, url):
    self.driver.get(url)
    time.sleep(10)
    self.closePopup()
    for ordering, data in self.scrape_sortings():
        for row in pd.DataFrame(data).iterrows():
            print(row)
            temp_dict = row[1].to_dict()
            temp_dict[self.fieldnames[13]] = ordering
            self.writer.writerow(temp_dict)
```

```python
scraper = KayakScraper()
scraper.scrape_flight(scraper.makeFlightURL("NYC","BOS","2019-09-06","2019-09-10"))
```

Using web scrapping with selenium, extract the following features:

To Airport · Price · From #stops · From Duration · From Airport · To Time · To Duration · To Carrier · From Time · From Carrier · To #stops · To Airport Stops · Sort · From Airport Stops

price	To_#stops	From_#stops	To_airport_stops	From_airport_stops	To_duration	From_duration	To_airport	From_airport	To_time	From_time	To_carrier	From_carrier	sort
$401	0	1	NaN	DTW	5h 49m	6h 54m	EWR - LAX	LAX - LGA	9:00 am — 11:49 am	9:50 pm — 7:44 am +1	United Airlines	Spirit Airlines	best
$472	0	0	NaN	NaN	5h 45m	5h 00m	EWR - LAX	LAX - EWR	7:00 am — 9:45 am	9:15 pm — 5:15 am +1	United Airlines	United Airlines	best
$472	0	0	NaN	NaN	5h 45m	5h 01m	EWR - LAX	LAX - EWR	7:00 am — 9:45 am	11:15 pm — 7:16 am +1	United Airlines	United Airlines	best
$472	0	0	NaN	NaN	5h 49m	5h 00m	EWR - LAX	LAX - EWR	9:00 am — 11:49 am	9:15 pm — 5:15 am +1	United Airlines	United Airlines	best
$472	0	0	NaN	NaN	5h 49m	5h 01m	EWR - LAX	LAX - EWR	9:00 am — 11:49	11:15 pm — 7:16 am +1	United Airlines	United Airlines	best

Work with the following features:

To Airport   From Airport

Price

To Carrier   From Carrier

To Time   From Time

To #stops   From #Stops

To Duration   From Duration
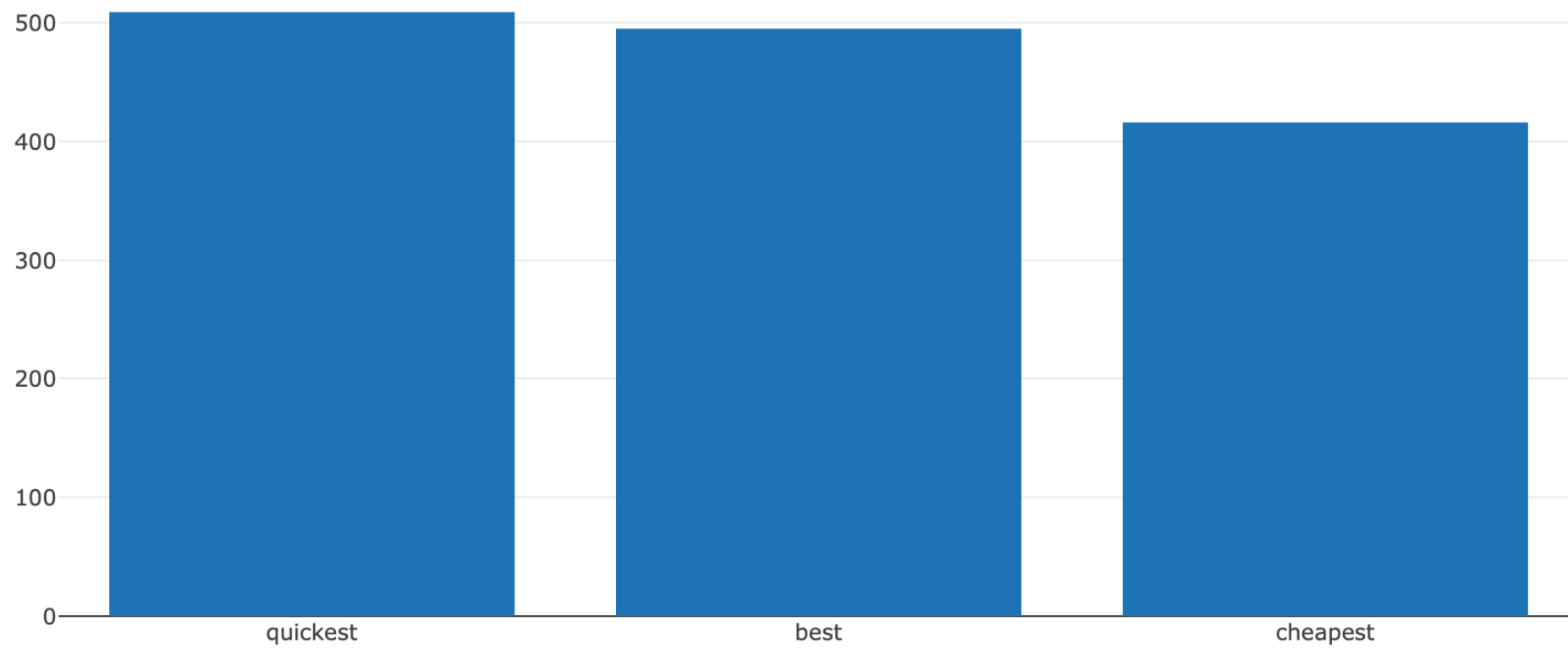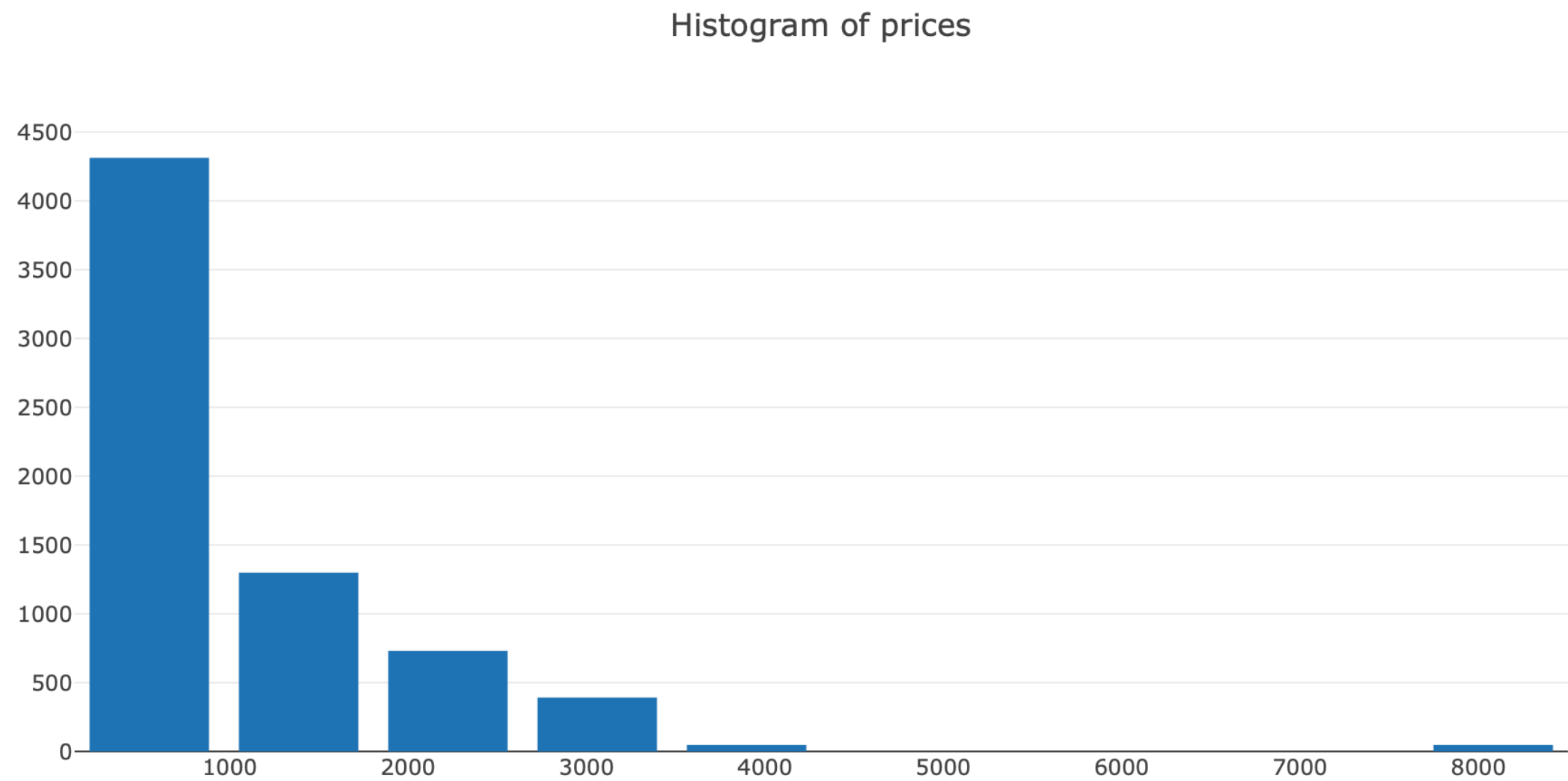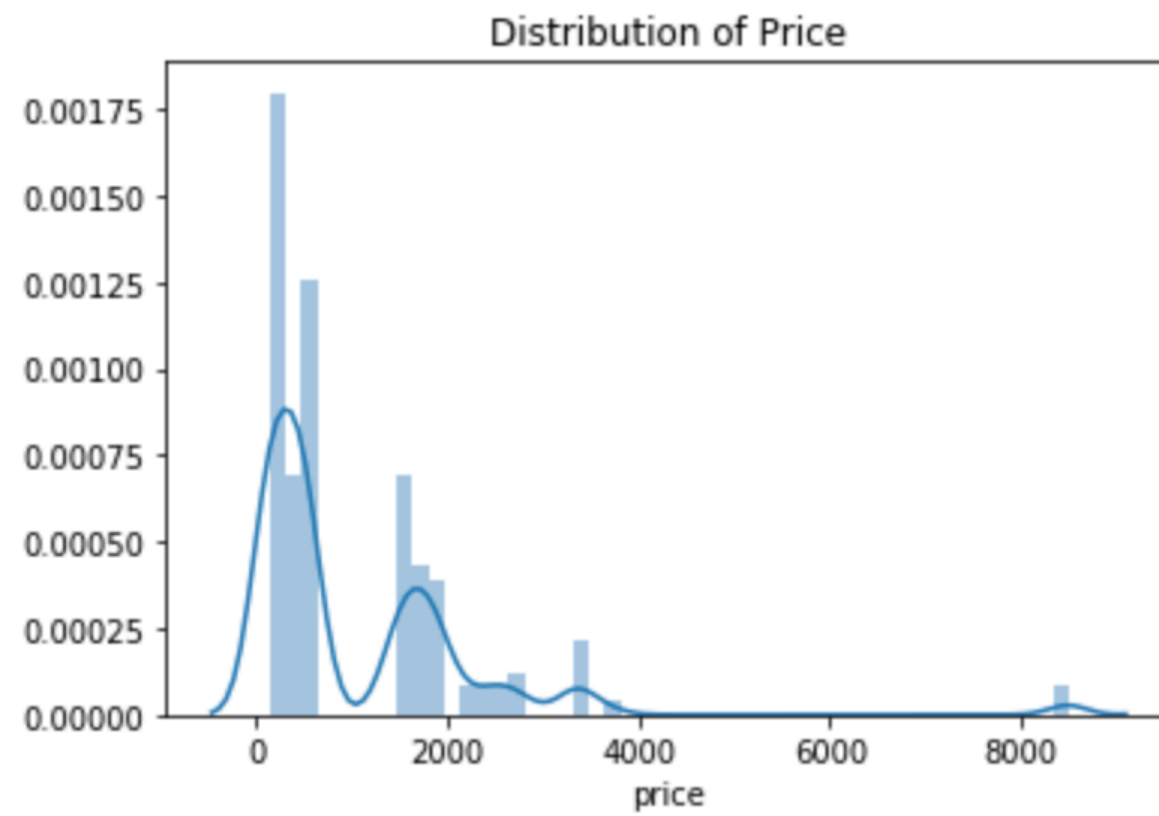
Sort

Carrier

Total mins

Total #stops

Sort
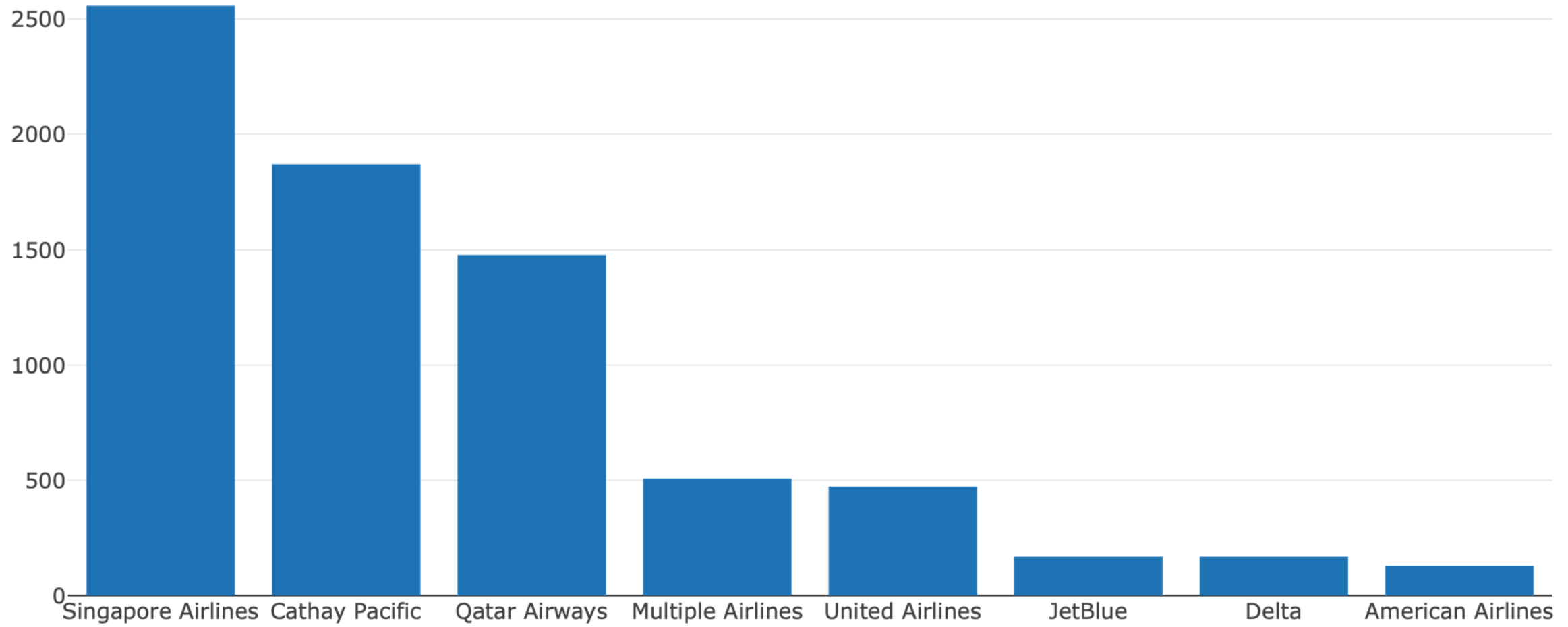
	price	To_airport	From_airport	To_time	From_time	sort	#stops	Total_mins	carrier
0	401	EWR - LAX	LAX - LGA	9:00 am — 11:49 am	9:50 pm — 7:44 am	best	1	763	Multiple Airlines
1	472	EWR - LAX	LAX - EWR	7:00 am — 9:45 am	9:15 pm — 5:15 am	best	0	645	United Airlines
2	472	EWR - LAX	LAX - EWR	7:00 am — 9:45 am	11:15 pm — 7:16 am	best	0	646	United Airlines
3	472	EWR - LAX	LAX - EWR	9:00 am — 11:49 am	9:15 pm — 5:15 am	best	0	649	United Airlines
4	472	EWR - LAX	LAX - EWR	9:00 am — 11:49 am	11:15 pm — 7:16 am	best	0	650	United Airlines

Median of Prices by Sort

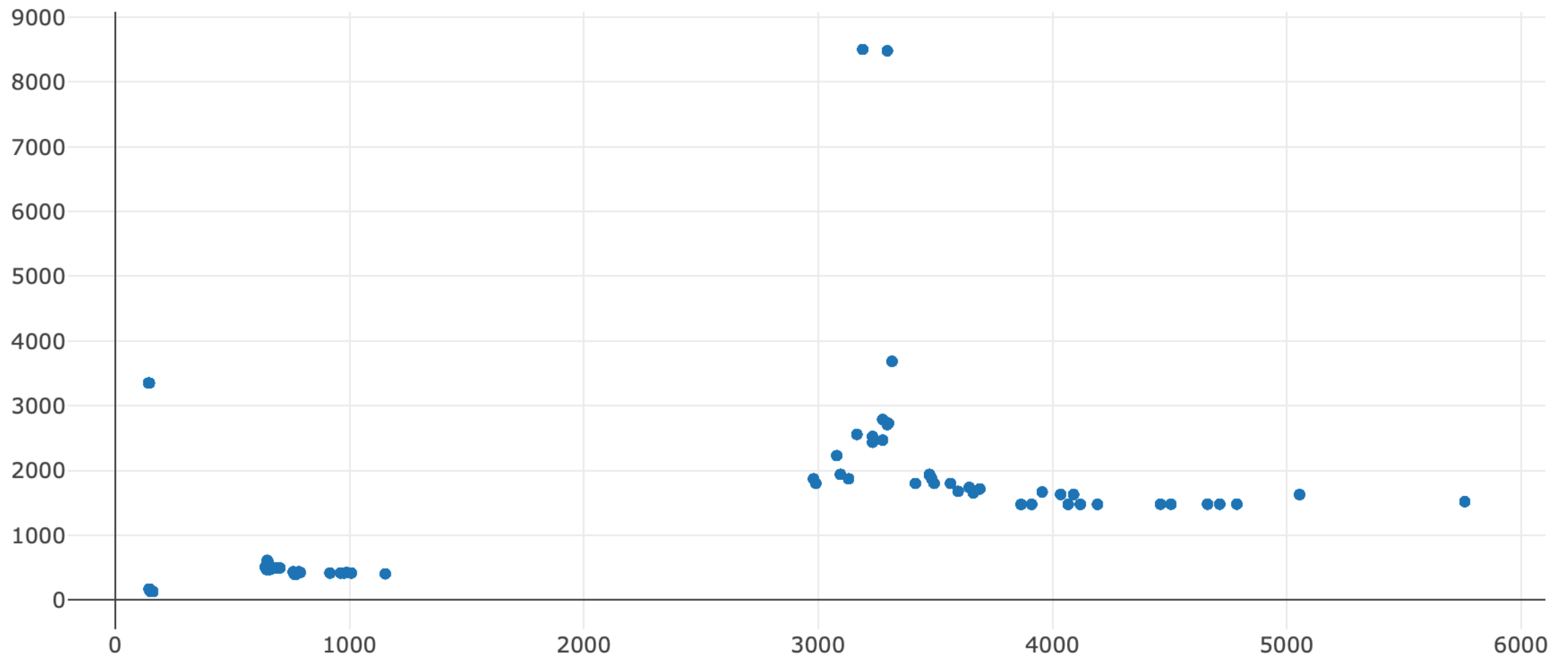Distribution of Price

Histogram of prices

Distribution of price by Airlines

Total mins versus price

Number of Stops versus price