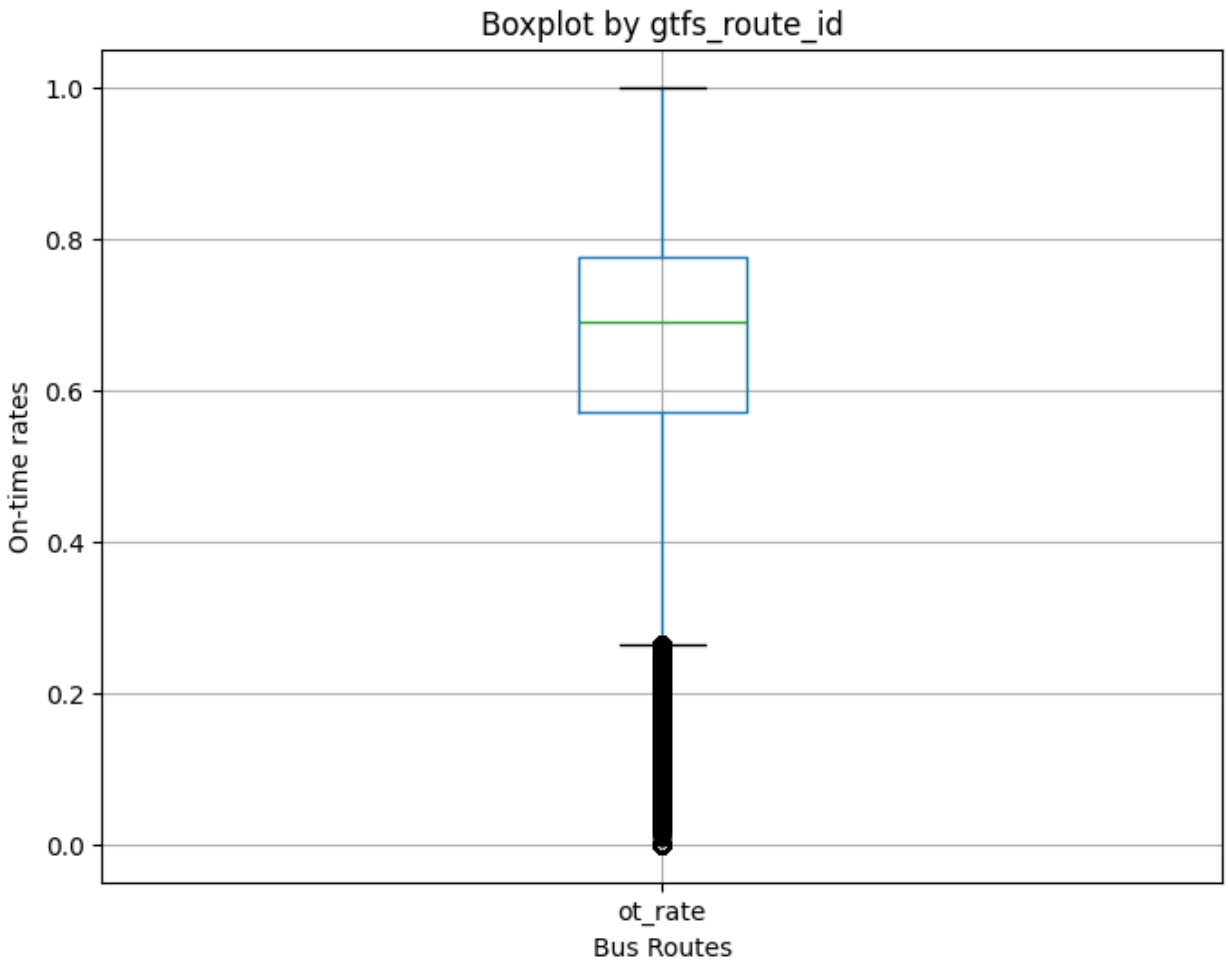
Are any bus lines more delayed than other lines?

We calculated the rate of on-time (ot_rate) by dividing the otp_denominator from the otp_nominator.

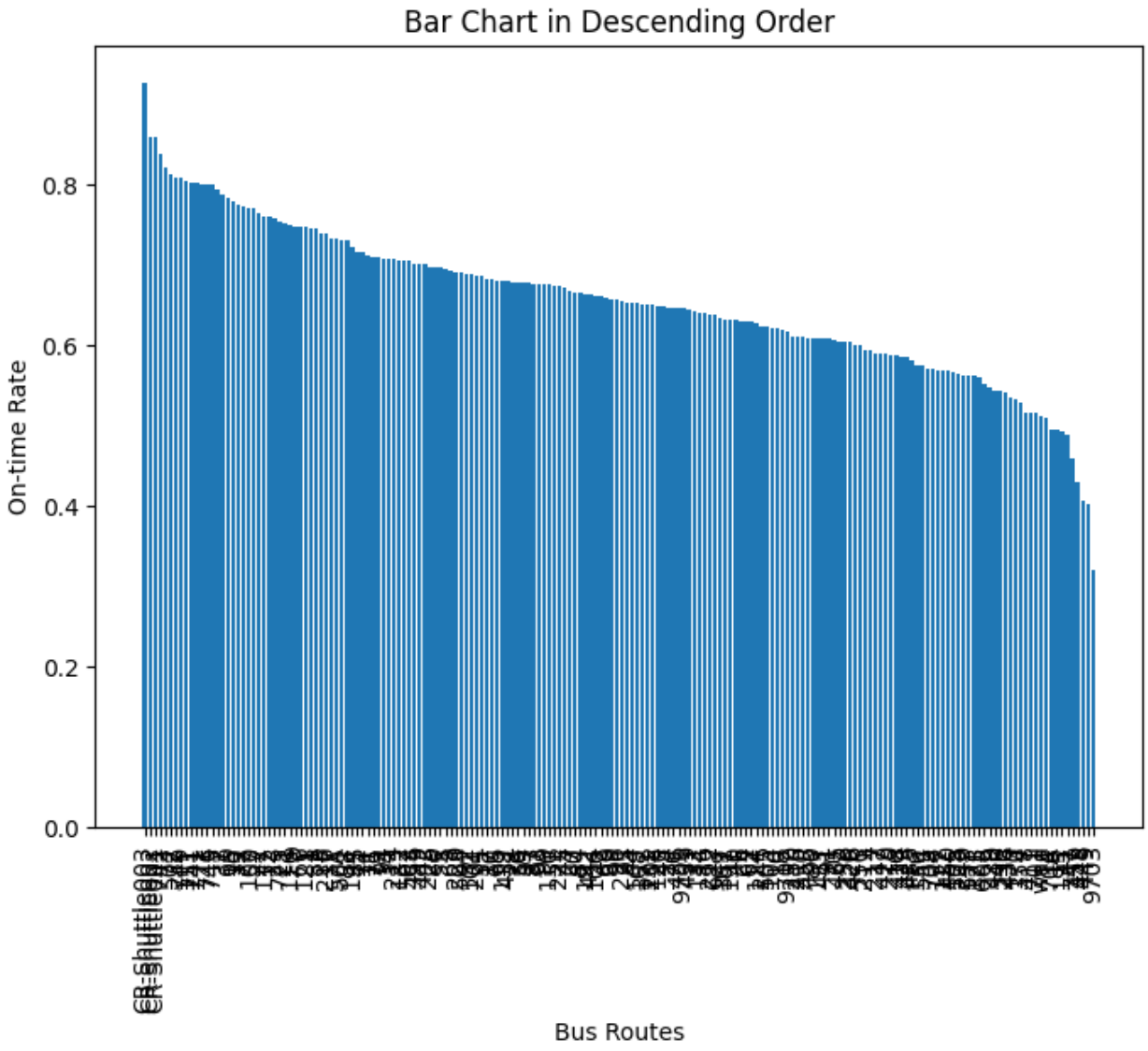
We are able to see the max, min, mean and median osdf bus route on-time rates. A rate of 0.0 indicates that the bus route was never on time. A rate of 1.0 means that the bus route was never late.

```
print ("median of bus on-time rate is ", df["ot_rate"].median())
print ("mean of bus on-time rate is", df["ot_rate"].mean() )
print ("max of bus on-time rate is", df["ot_rate"].max() )
print ("min of bus on-time rate is", df["ot_rate"].min() )
```

```
median of bus on-time rate is  0.689119170984456
mean of bus on-time rate is 0.6637900441846418
max of bus on-time rate is 1.0
min of bus on-time rate is 0.0
```



By grouping all bus routes (gtfs_route_id) and taking the mean of each bus route's on-time rates, we then plotted a boxplot of the on-time rate. Although many outliers have low on-time rate, the majority bus routes can achieve on-time rates between 0.5 and 0.8 with a median of 0.69.



Then we sorted the bus routes by their on-time rate with descending order. Then we retrieved the Top 10 and Bottom 10 bus routes with the highest on-time rates and the lowet on-time rates. Bus route CR-Shuttle003, CR-Shuttle002 and CR-Shuttle001 have the highest on-time rate. Bus routes 448, 449 and 9703 have the lowest on-time rate.

	gtfs_route_id	ot_rate
182	CR-Shuttle003	0.925859
181	CR-Shuttle002	0.858203
180	CR-Shuttle001	0.858203
147	742	0.837185
144	73	0.820220
112	502	0.813195
65	32	0.807782
151	749	0.807251
11	111	0.803600
153	751	0.801902

	gtfs_route_id	ot_rate
24	14	0.509825
140	70A	0.494182
31	19	0.493452
36	195	0.491992
82	41	0.488934
150	747	0.458202
106	459	0.429970
99	448	0.406302
100	449	0.402552
178	9703	0.320094