

## Working With Parquet Format

TLC is switching to the Parquet file type for storing raw trip data on our website. Parquet is the industry standard for working with big data. Using Parquet format results in reduced file sizes and increased speeds. However, we have been using the CSV format for a while and the Parquet format might be new to some users.

Below are examples of how to open Parquet files using R and Python:

### Python:

```
import pyarrow.parquet as pq
trips = pq.read_table('trips.parquet')
trips = trips.to_pandas()
```

Additional documentation: <https://arrow.apache.org/docs/python/parquet.html>

### R:

```
library(arrow)
trips <- read_parquet('trips.parquet')
```

Additional documentation: [https://arrow.apache.org/docs/r/reference/read\\_parquet.html](https://arrow.apache.org/docs/r/reference/read_parquet.html)