

Appendix A

Dataset Sizes Before and After Cleaning

Month	Raw Data		Process #1 (Excel)		Process #2 (BigQuery)	
	Columns	Rows	Columns	Rows	Columns	Rows
September 2024	13	821,277	8	537,235	8	502,790
October 2024	13	616,282	8	449,115	8	418,110
November 2024	13	335,076	8	245,972	8	225,306
December 2024	13	178,373	8	130,731	8	118,643
January 2025	13	138,690	8	101,241	8	90,494
February 2025	13	151,881	8	109,570	8	98,068
March 2025	13	298,156	8	209,841	8	191,613
April 2025	13	371,342	8	258,857	8	237,147
May 2025	13	502,457	8	340,635	8	315,941
June 2025	13	678,905	8	446,364	8	419,995
July 2025	13	763,433	8	494,294	8	463,666
August 2025	13	790,178	8	515,534	8	483,980

Feature Engineering

New Column Name	Description	Data Type	Creation Details
ride_day	Day of the week on which the ride took place	String	=TEXT(<i>ref</i> , "dddd"), where <i>ref</i> are the records in the started_at column.
ride_length_sec	The length of ride in seconds	Float	=(ended_at_ <i>ref</i> - started_at_ <i>ref</i>)*86400, where <i>ref</i> are cells on the same row number, e.g., C2 and D2.
ride_length_min	The length of ride in minutes	Float	=(ended_at_ <i>ref</i> - started_at_ <i>ref</i>)*1440, where <i>ref</i> are cells on the same row number, e.g., C2 and D2.

SQL Queries

```
CREATE TABLE `sk-repo.cyclistic.sep_24_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.sep_24_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.oct_24_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.oct_24_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.nov_24_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.nov_24_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.dec_24_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.dec_24_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.jan_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.jan_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.feb_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.feb_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.mar_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.mar_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.apr_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.apr_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.may_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.may_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.jun_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.jun_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.jul_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.jul_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

```
CREATE TABLE `sk-repo.cyclistic.aug_25_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.aug_25_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```