# MAVEN TAXI CHALLENGE

DATA CLEANSING STEPS:

|  |  |
| --- | --- |
| 1. *Let’s stick to trips that were NOT sent via “store and forward”* | Filtered **store\_and\_fwd\_flag** for *Yes (Y)* |
| 1. *I’m only interested in street-hailed trips paid by card or cash, with a standard rate* | Filtered **trip\_type = 1= Street-hail (1)**  Filtered **RatecodeID = 1 = Standard rate**  Filtered **payment\_type = 1= Credit card; 2= Cash** |
| 1. *We can remove any trips with dates before 2017 or after 2020, along with any trips with pickups or drop-offs into unknown zones* |  |
| 1. *Let’s assume any trips with no recorded passengers had 1 passenger* | Changed **passenger\_count = 0 to 1** |
| 1. *If a pickup date/time is AFTER the drop-off date/time, let’s swap them* |  |
| 1. *We can remove trips lasting longer than a day, and any trips which show both a distance and fare amount of 0* | Removed (**trip\_distance = 0** & **fare = 0)**  Still have trip\_distance = 0 but fare amounts present (Solved in 9) |
| 1. *If you notice any records where the fare, taxes, and surcharges are ALL negative, please make them positive* | Fare\_amount,  Extra,  Mta\_tax |
| 1. *For any trips that have a fare amount but have a trip distance of 0, calculate the distance this way: (Fare amount - 2.5) / 2.5* | Done |
| 1. *For any trips that have a trip distance but have a fare amount of 0, calculate the fare amount this way: 2.5 + (trip distance x 2.5)* | Done |
| 1. *Some trip distances were 0 but we still have trip amount* |  |

* Date in 454 calendar table starts from 05-02-2017 ( Feb 5) but pick up date from 01-01-2017

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |