Data Preprocessing

| **Case** | **Column** | **Method** |
| --- | --- | --- |
| Duplicated Data | partitionKey | Remove the duplicated  simultaneous looking and bookings from the same user:  1. More than one booking completed.  2. A prior order is completed while other orders are not. |
| Missing Data | startAddress | Remove the users who do not finish the booking in the LookUp Subset (lk\_subset).  Fill in the missing address with latitude and longitude in the Booking Subset (bk\_subset). |
| endAddress |
| DropOff\_Latitude | Remove the value missing rows. |
| DropOff\_Longitude |
| Feature Engineering | StartState | Matching states to each row by the booking address. |
| EndState |
| createdAt | Transfer the data type to datetime format. |
| BookingCreatedTime |
| DispatchStatus | Transfer NotCompleted=0, Completed=1. |
| BookingCreatedHour | Extract the exact Hour, Minute and Day in BookingCreatedTime. |
| BookingCreatedMin |
| BookingCreatedDay |