# Cleaning steps taken for cyclistic data 202405

|  |  |  |
| --- | --- | --- |
| Data column | Principle | Action and outcome |
| Ride\_id | Ensure no blank values | The Sort feature showed no blanks, but it did show 400 rides that required scientific notation, going as large as having 300 numbers in one id. These were not removed, however analysis concerning ride\_id should not be performed. |
|  | Ensure no duplicate values as each id should be unique. | Remove duplicate function used, 9 duplicates removed. |
| Rideable\_type | Ensure only classic\_bike, docked\_bike, electric bike were values. | Pivot table showed this to be true |
|  |  |  |
| Started\_at | Ensure no blank values | Sort feature – no values before the start of the month or after end. |
|  |  |  |
| Ended\_at | Ensure no blank values | Sort feature - no values before the start of the month, 2 values after end of the month. Not removed as they started before end. |
|  |  |  |
| Start\_station\_name | Ensure no blank values | Pivot table showed that there were 109,048 blank spaces. Vlookup was used to populate based off coordinates, but these coordinates did not corelate to any known stations. The rows were left but any future analysis on stations needs to consider this limitation. |
|  |  |  |
| start\_station\_id | Ensure no blank values | Pivot table showed that there were 109,048 blank spaces. Vlookup was used to populate based off coordinates, but these coordinates did not corelate to any known stations. The rows were left but any future analysis on stations needs to consider this limitation. |
|  |  |  |
| end\_station\_name | Ensure no blank values | Pivot table showed that there were 112,731 blank spaces. Vlookup was used to populate based off coordinates, but these coordinates did not corelate to any known stations. The rows were left but any future analysis on stations needs to consider this limitation. |
|  |  |  |
| end\_station\_id | Ensure no blank values | Pivot table showed that there were 112,731 blank spaces. Vlookup was used to populate based off coordinates but these coordinates did not corelate to any known stations. The rows were left but any future analysis on stations needs to consider this limitation. |
|  |  |  |
| Start\_lat | Ensure no blank values | Filter feature. There are no blank values in these fields. |
|  |  |  |
| Start\_lng | Ensure no blank values | Filter feature. There are no blank values in these fields. |
|  |  |  |
| end\_lat | Ensure no blank values | Filter feature. There are 783 blank values in these fields. All of these are missing all data on ending stations or locations. |
|  |  |  |
| End\_lng | Ensure no blank values | Filter feature. There are 783 blank values in these fields. All of these are missing all data on ending stations or locations. |
|  |  |  |
| Member\_casual | Ensure no blank values | Pivot table. Only causal and member were in these tables, no blanks. |
| ride\_length | Values outside of range | Note that 270 values are negative or zero. Due to the small number of these values they were not removed. |