**Data Quality Issues:**

* The most difficult and time-consuming issue faced in the data was translating the language from Spanish to English (As the team didn’t have any Spanish speakers).
* The Google Form had some questions which had checkboxes for answers which allowed the user to input multiple answers where the response was intended to be only one.
* The data had many missing values. There were some questions on the Google Form which were not mandatory. This could be one of the reasons for multiple missing values.
* These forms were filled manually hence this can induce human error or biases in the data collected.
* Dropping columns – The columns important for the analysis (The General Public and Press) is only considered.
* Unable to clean data row wise – The rows indicated duplicate values, for example the Hospital BOL000 often showed double entries for every reporting week with different values.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country Name | (HC)# | Hospital code | Entries (#) | Type | (Type #) | Administrative code |
| Apure | 1 | APU000 | 28 | III | 23 | MPPS, Gob. Estatal |
|  |  |  |  | IV | 5 | MPPS, INSALAD, INSULAD APURE |
| Bolivar | 2 | BOL000 | 51 | III | 20 | IVSS |
|  |  |  |  | IV | 31 | ISP  ISP Bolivar,  ISP Estado Bolivar |
| DTTO Capital | 10 | DCA001 | 9 | III | 1 | MPPS |
|  |  |  |  | IV | 8 | MPPS, Gob. Estatal |
|  |  | DCA005 | 39 | II | 1 | IVSS |
|  |  |  |  | IV | 38 | IVSS |
| Guarico | 1 | GUA000 | 38 | III | 33 | MPSS |
|  |  |  |  | IV | 5 | MPSS |
| Portuguesa | 1 | POR000 | 35 | III | 34 | MPPS, Gob. Estatal |
|  |  |  |  | IV | 1 | MPPS, Gob. Estatal |
| Vargas | 1 | VAR000 | 37 | II | 15 | IVSS |
|  |  |  |  | III | 22 | IVSS |

**Cleaning the Data:**

* Track 1 was chosen for analysis and the Metrics was chosen carefully to cater to the audience (General Public and Press)
* The columns were converted manually from Spanish to English using ‘.replace()’ in Python.
* Missing values was filled if there were columns indicating a similarity. For example, columns like Renal replacement therapy availability and Renal replacement therapy operability are similar, where missing values can be filled in one column based on the value in the previous column.
* The questions that had a ‘yes’ (‘Si’) or ‘No’ answer was converted to 1 and 0 respectively.

**Data Wrangling:**

* The metrics important to cater to the audience was identified. These includes the region, hospital code, if the hospital is approachable for an emergency or not, etc.
* The hospitals were considered in two different categories such as emergency services and regular services. The metric considered for the analysis is based on the basic requirement a hospital must have such as emergency supplies and ability to perform minor surgeries such as Appendicitis for emergency cases.

**REFERENCE:**

FOR SUPPORTING THE IDEA OF HAVING EMERGENCY KITS .

<https://www.aclsmedicaltraining.com/crash-carts/>