1. **INTRODUCTION**

This dataset can be found on Kaggle, under the UNCOVER COVID-19 Challenge dataset (<https://www.kaggle.com/roche-data-science-coalition/uncover/tasks>). The UNCOVER COVID-19 Challenge dataset consists of many different datasets from a wide variety of sources. Under the Canada\_Hosp1\_COVID\_InpatientData.xlsx file (containing data about patients at a Canadian hospital), there is a sheet named Data-at-admission (the patients’ data recorded upon admission), which we separated out and saved as Canada\_Hosp1\_COVID\_Data\_At\_Admission.csv. We chose this dataset due to its high number of variables (54 columns) and decent mix of categorical and continuous variables.

1. **DATA SET DESCRIPTION**

After cleaning, this dataset contains 508 samples with 40 columns with various data types. A complete listing is shown in Table 1. Before cleaning, the dataset contained 508 samples with 55 columns with various data types. We decided that any column with less than 40 non-null values were to discarded immediately, leading to us dropping columns *ethnicity\_other, received\_covid\_vaccine, covid\_vaccine, pao2, pao2\_fio2, ph, high\_senstivity\_cardiac\_troponin, esr, ferritin,* and *hs\_crp*. *id* was dropped as it was (for our purposes) simply an alternate index. *ethnicity* was dropped as its data was apparently scrubbed, consisting entirely of “[]”. *aptt\_aptr* was dropped as it consisted entirely of numbers but used the string data type, had only 288 non-null values, and did not seem useful after some research as to what it meant. *meds* was dropped due to having too many unique values to be of use, along with *comorb\_other.*

**Table 1: Data Types and Missing Data**

| *Variable Name* | *Data Type* | *Missing Data (%)* |
| --- | --- | --- |
| reason | nominal/object | 0% |
| age | ratio/int64 | 0% |
| sex | nominal/object | 0% |
| height | ratio/float64 | 46.46% |
| weight | ratio/float64 | 29.53% |
| comorb | nominal/object | 25.39% |
| smoke\_hist | nominal/object | 41.73% |
| year\_quit | interval/float64 | 91.54% |
| er\_2\_weeks | nominal/object | 0% |
| admission\_disposition | nominal/object | 0% |
| systolic\_bp | ratio/int64 | 0% |
| diastolic\_bp | ratio/int64 | 0% |
| heart\_rate | ratio/int64 | 0% |
| respiratory\_rate | ratio/int64 | 0% |
| oxygen\_sat | ratio/int64 | 0% |
| temp | interval/float64 | 0% |
| motor | nominal/object | .39% |
| verbal | nominal/object | .39% |
| eye | nominal/object | .39% |
| intubated | nominal/object | 0% |
| wbc | ratio/float64 | .39% |
| rbc | ratio/float64 | .39% |
| hemoglobin | ratio/float64 | .39% |
| hematocrit | ratio/float64 | .39% |
| mcv | ratio/float64 | .39% |
| mch | ratio/float64 | .39% |
| mchc | ratio/float64 | .39% |
| rdw | ratio/float64 | .39% |
| platelet\_count | ratio/float64 | .39% |
| pt | ratio/float64 | 37.99% |
| alt | ratio/float64 | 20.67% |
| ast | ratio/float64 | 20.67% |
| serum\_creatinine | ratio/float64 | .39% |
| sodium | ratio/float64 | .39% |
| potassium | ratio/float64 | .39% |
| total\_serum\_bilirubin | ratio/float64 | 20.67% |
| lactate | ratio/float64 | 28.74% |
| inr | ratio/float64 | 37.99% |
| d\_dimer | ratio/float64 | 50.39% |
| crp | ratio/float64 | 72.05% |

1. **DATA SET SUMMARY STATISTICS**

This section consists of tables representing various statistical summaries of the dataset. Table 2 shows the basic statistics (count, mean, standard deviation, etc.) of each continuous variable. Tables 3A-J are summaries of the categorical variables, each table showing each value and its frequency/proportion for a given categorical variable. Tables 4A-E are correlation matrices, split up due to the large number of continuous variables in the dataset. Figure 0 is the heatmap of thee correlation matrix, a single visual representation of the data within Tables 4A-E.

**Table 2: Summary Statistics for Canada\_Hosp1\_COVID\_Data\_At\_Admission**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Variable Name* | *Count* | *Mean* | *STDev* | *Min* | *25th* | *50th* | *75th* | *Max* |
| age | 508 | 66 | 16.4 | 19 | 55.8 | 66 | 78 | 100 |
| height | 272 | 166.9 | 11 | 125 | 159 | 167 | 175 | 198 |
| weight | 358 | 80.3 | 21.8 | 27.7 | 65.8 | 76.7 | 89.8 | 199.6 |
| year\_quit | 43 | 1999.7 | 15.4 | 1971 | 1990 | 2000 | 2013.5 | 2021 |
| systolic\_bp | 508 | 129.4 | 22.1 | 85 | 115 | 127 | 143 | 227 |
| diastolic\_bp | 508 | 76 | 12.9 | 22 | 68 | 75 | 83 | 145 |
| heart\_rate | 508 | 97.4 | 17.5 | 41 | 86 | 98 | 109 | 161 |
| respiratory\_rate | 508 | 24.9 | 7.2 | 12 | 20 | 24 | 28 | 59 |
| oxygen\_sat | 508 | 93.1 | 6.1 | 47 | 92 | 94 | 96 | 100 |
| temp | 508 | 37.8 | .96 | 35.5 | 37 | 37.7 | 38.5 | 40.5 |
| wbc | 506 | 8 | 3.7 | 2 | 5.5 | 7.4 | 9.7 | 33.4 |
| rbc | 506 | 4.6 | .72 | 2.78 | 4.2 | 4.6 | 5 | 8.2 |
| hemoglobin | 506 | 129.9 | 18 | 74 | 119 | 131 | 142 | 180 |
| hematocrit | 506 | .395 | .05 | .23 | .36 | .397 | .43 | .55 |
| mcv | 506 | 86 | 7.4 | 56.7 | 83 | 86.5 | 90.3 | 112.2 |
| mch | 506 | 28.3 | 2.8 | 17.3 | 27 | 28.7 | 30 | 38.2 |
| mchc | 506 | 328.6 | 13.5 | 272 | 319 | 330 | 338 | 382 |
| rdw | 506 | 14 | 1.9 | 10.3 | 12.8 | 13.6 | 14.6 | 24.8 |
| platelet\_count | 506 | 230.9 | 98 | 19 | 163 | 214 | 276 | 671 |
| pt | 315 | 14.6 | 5.4 | 11.7 | 13 | 13.5 | 14.4 | 72.6 |
| alt | 403 | 51.2 | 51 | 6 | 25 | 38 | 59 | 586 |
| ast | 403 | 61.4 | 78.5 | 11 | 33 | 47 | 68 | 1307 |
| serum\_creatinine | 506 | 114.8 | 92 | 30 | 76 | 95 | 121 | 1179 |
| sodium | 506 | 135.3 | 5.2 | 112 | 133 | 135 | 138 | 161 |
| potassium | 506 | 3.89 | .53 | 2.1 | 3.5 | 3.9 | 4.2 | 6.3 |
| total\_serum\_bilirubin | 403 | 10.6 | 5.7 | 3 | 7 | 9 | 12 | 40 |
| lactate | 362 | 2.33 | 1.25 | .6 | 1.6 | 2 | 2.7 | 13.8 |
| inr | 315 | 1.16 | .68 | .9 | 1 | 1 | 1.1 | 8.9 |
| d\_dimer | 252 | 1346.4 | 898.4 | 269 | 694 | 1030 | 1764 | 4001 |
| crp | 142 | 109.1 | 74.9 | 3 | 53 | 93.2 | 158.8 | 371 |

Table 3.A: Proportions for ‘reason’ (n=508)

|  |  |  |
| --- | --- | --- |
| *Category* | *Freq.* | *Proportion* |
| Acute respiratory distress syndrome | 2 | 0.4% |
| COVID-19 | 220 | 43.3% |
| Coronavirus infection | 2 | 0.4% |
| Cough | 2 | .4% |
| Febrile respiratory illness | 7 | 1.4% |
| Fever | 22 | 4.3% |
| Hypoxemia | 22 | 4.3% |
| Myalgia | 1 | 0.2% |
| Pneumonia | 136 | 26.8% |
| Pneumonia due to COVID-19 | 41 | 8.1% |
| Probable COVID-19 | 8 | 1.6% |
| Respiratory failure | 7 | 1.4% |
| Respiratory tract infection | 1 | 0.2% |
| Shortness of breath | 34 | 6.7% |
| Shortness of breath with exposure to COVID-19 | 1 | 0.2% |
| Tachypnea | 1 | 0.2% |
| Viral pneumonia | 1 | 0.2% |

Table 3.B: Proportions for ‘sex’ (n=508)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| Female | 212 | 41.7% |
| Male | 296 | 58.3% |

Table 3.C: Proportions for ‘smoke\_hist’ (n=296)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| Ex-smoker | 66 | 22.3% |
| Non-smoker | 218 | 73.6% |
| Smoker | 12 | 4.1% |

Table 3.D: Proportions for ‘er\_2\_weeks’ (n=508)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| No | 43 | 8.5% |
| Yes | 465 | 91.5% |

Table 3.E: Proportions for ‘admission\_disposition’ (n=508)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| ICU | 43 | 8.5% |
| WARD | 465 | 91.5% |

Table 3.F: Proportions for ‘comorb’ (n=379)

| *Category* | *Freq.* | *Prop.* |
| --- | --- | --- |
| Asthma | 17 | 4.5% |
| Asthma, Chronic cardiac disease | 19 | 5.0% |
| Asthma, Chronic cardiac disease, Chronic pulmonary disease | 1 | 0.3% |
| Asthma, Chronic cardiac disease, Chronic pulmonary disease, Diabetes | 3 | 0.8% |
| Asthma, Chronic cardiac disease, Diabetes | 12 | 3.2% |
| Asthma, Chronic pulmonary disease | 1 | 0.3% |
| Asthma, Diabetes | 1 | 0.3% |
| Chronic cardiac disease | 152 | 40.1% |
| Chronic cardiac disease, Chronic pulmonary disease | 9 | 2.4% |
| Chronic cardiac disease, Chronic pulmonary disease, Diabetes | 7 | 1.8% |
| Chronic cardiac disease, Diabetes | 125 | 33.0% |
| Chronic pulmonary disease | 5 | 1.3% |
| Chronic pulmonary disease, Diabetes | 2 | 0.5% |
| Diabetes | 25 | 6.6% |

Table 3.G: Proportions for ‘motor’ (n=506)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| 1 | 2 | 0.4% |
| 2 | 1 | 0.2% |
| 3 | 2 | 0.4% |
| 4 | 2 | 0.4% |
| 5 | 20 | 4.0% |
| 6 | 479 | 94.7% |

Table 3.H: Proportions for ‘verbal’ (n=506)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| 1 | 9 | 1.8% |
| 2 | 10 | 2.0% |
| 3 | 2 | 0.4% |
| 4 | 47 | 9.3% |
| 5 | 438 | 86.6% |

Table 3.I: Proportions for ‘eye’ (n=506)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| 1 | 3 | 0.6% |
| 2 | 3 | 0.6% |
| 3 | 13 | 2.6% |
| 4 | 487 | 96.2% |

Table 3.J: Proportions for ‘intubated’ (n=508)

|  |  |  |
| --- | --- | --- |
| *Category* | *Frequency* | *Proportion (%)* |
| No | 493 | 97.0% |
| Yes | 15 | 3.0% |

Table 4: Correlation Table/Tables

| Table 4A | *age* | *height* | *weight* | *year\_quit* | *systolic\_bp* | *diastolic\_bp* |
| --- | --- | --- | --- | --- | --- | --- |
| *age* | 1.00 | -0.27 | -0.38 | -0.34 | 0.07 | -0.17 |
| *height* | -0.27 | 1.00 | 0.47 | -0.02 | 0.02 | 0.10 |
| *weight* | -0.38 | 0.47 | 1.00 | -0.10 | -0.01 | 0.08 |
| *year\_quit* | -0.34 | -0.02 | -0.10 | 1.00 | 0.08 | 0.24 |
| *systolic\_bp* | 0.07 | 0.02 | -0.01 | 0.08 | 1.00 | 0.57 |
| *diastolic\_bp* | -0.17 | 0.10 | 0.08 | 0.24 | 0.57 | 1.00 |
| *heart\_rate* | -0.35 | 0.14 | 0.10 | 0.17 | 0.13 | 0.29 |
| *respiratory\_rate* | -0.15 | 0.01 | 0.08 | 0.16 | 0.11 | 0.12 |
| *oxygen\_sat* | -0.03 | 0.01 | -0.09 | -0.38 | 0.02 | 0.10 |
| *temp* | -0.02 | 0.05 | 0.05 | 0.00 | 0.09 | 0.01 |
| *wbc* | 0.02 | -0.07 | 0.00 | -0.03 | 0.04 | 0.02 |
| *rbc* | -0.29 | 0.31 | 0.22 | 0.23 | 0.02 | 0.19 |
| *hemoglobin* | -0.18 | 0.32 | 0.16 | 0.25 | -0.04 | 0.15 |
| *hematocrit* | -0.16 | 0.31 | 0.20 | 0.17 | -0.05 | 0.16 |
| *mcv* | 0.29 | -0.04 | -0.08 | -0.20 | -0.09 | -0.10 |
| *mch* | 0.20 | 0.01 | -0.10 | -0.01 | -0.07 | -0.09 |
| *mchc* | -0.09 | 0.11 | -0.09 | 0.21 | 0.03 | 0.00 |
| *rdw* | 0.12 | -0.09 | 0.08 | -0.11 | 0.00 | 0.01 |
| *platelet\_count* | -0.12 | -0.10 | 0.04 | -0.33 | -0.01 | 0.03 |
| *pt* | 0.12 | 0.20 | 0.03 | -0.18 | -0.02 | -0.03 |
| *alt* | -0.16 | 0.12 | 0.00 | -0.01 | 0.00 | 0.04 |
| *ast* | -0.01 | 0.10 | -0.03 | 0.05 | 0.00 | -0.01 |
| *serum\_creatinine* | 0.16 | 0.04 | -0.01 | -0.05 | 0.02 | -0.04 |
| *sodium* | 0.17 | 0.00 | -0.06 | -0.15 | -0.11 | -0.01 |
| *potassium* | 0.19 | 0.14 | 0.00 | 0.02 | 0.03 | 0.04 |
| *total\_serum\_bilirubin* | 0.13 | 0.20 | 0.05 | 0.10 | 0.01 | 0.03 |
| *lactate* | 0.12 | 0.12 | 0.04 | 0.22 | 0.04 | 0.05 |
| *inr* | 0.11 | 0.20 | 0.02 | -0.17 | -0.01 | -0.02 |
| *d\_dimer* | 0.31 | -0.16 | -0.22 | -0.68 | -0.03 | -0.04 |
| *crp* | -0.10 | 0.10 | 0.27 | -0.49 | -0.01 | -0.16 |

| Table 4B | *heart\_rate* | *respiratory\_rate* | *oxygen\_sat* | *temp* | *wbc* | *rbc* |
| --- | --- | --- | --- | --- | --- | --- |
| *age* | -0.35 | -0.15 | -0.03 | -0.02 | 0.02 | -0.29 |
| *height* | 0.14 | 0.01 | 0.01 | 0.05 | -0.07 | 0.31 |
| *weight* | 0.10 | 0.08 | -0.09 | 0.05 | 0.00 | 0.22 |
| *year\_quit* | 0.17 | 0.16 | -0.38 | 0.00 | -0.03 | 0.23 |
| *systolic\_bp* | 0.13 | 0.11 | 0.02 | 0.09 | 0.04 | 0.02 |
| *diastolic\_bp* | 0.29 | 0.12 | 0.10 | 0.01 | 0.02 | 0.19 |
| *heart\_rate* | 1.00 | 0.28 | -0.11 | 0.21 | 0.17 | 0.22 |
| *respiratory\_rate* | 0.28 | 1.00 | -0.29 | 0.12 | 0.24 | 0.09 |
| *oxygen\_sat* | -0.11 | -0.29 | 1.00 | 0.03 | -0.14 | -0.03 |
| *temp* | 0.21 | 0.12 | 0.03 | 1.00 | -0.01 | -0.01 |
| *wbc* | 0.17 | 0.24 | -0.14 | -0.01 | 1.00 | -0.03 |
| *rbc* | 0.22 | 0.09 | -0.03 | -0.01 | -0.03 | 1.00 |
| *hemoglobin* | 0.16 | 0.02 | -0.05 | -0.04 | -0.01 | 0.73 |
| *hematocrit* | 0.16 | 0.02 | -0.03 | -0.05 | -0.02 | 0.80 |
| *mcv* | -0.16 | -0.14 | 0.03 | -0.05 | 0.01 | -0.53 |
| *mch* | -0.12 | -0.12 | 0.00 | -0.03 | 0.03 | -0.47 |
| *mchc* | 0.03 | 0.00 | -0.06 | 0.03 | 0.05 | -0.07 |
| *rdw* | 0.04 | 0.05 | -0.06 | 0.09 | 0.08 | 0.08 |
| *platelet\_count* | 0.10 | 0.09 | -0.05 | -0.16 | 0.36 | -0.02 |
| *pt* | -0.05 | 0.00 | -0.06 | -0.07 | -0.02 | -0.13 |
| *alt* | 0.10 | 0.10 | -0.01 | -0.01 | 0.06 | 0.06 |
| *ast* | 0.02 | 0.07 | -0.08 | -0.05 | 0.02 | 0.00 |
| *serum\_creatinine* | -0.11 | -0.05 | 0.04 | 0.04 | 0.03 | -0.14 |
| *sodium* | -0.01 | -0.02 | 0.07 | -0.05 | 0.02 | 0.00 |
| *potassium* | -0.11 | -0.07 | 0.07 | -0.11 | -0.01 | -0.08 |
| *total\_serum\_bilirubin* | -0.01 | 0.13 | -0.11 | -0.08 | 0.18 | 0.09 |
| *lactate* | 0.15 | 0.24 | -0.31 | -0.16 | 0.25 | 0.16 |
| *inr* | -0.05 | 0.01 | -0.04 | -0.08 | -0.03 | -0.13 |
| *d\_dimer* | 0.05 | 0.16 | -0.11 | -0.02 | 0.17 | -0.27 |
| *crp* | 0.06 | 0.40 | -0.29 | -0.09 | 0.44 | -0.03 |

| Table 4C | *hemoglobin* | *hematocrit* | *mcv* | *mch* | *mchc* | *rdw* |
| --- | --- | --- | --- | --- | --- | --- |
| *age* | -0.18 | -0.16 | 0.29 | 0.20 | -0.09 | 0.12 |
| *height* | 0.32 | 0.31 | -0.04 | 0.01 | 0.11 | -0.09 |
| *weight* | 0.16 | 0.20 | -0.08 | -0.10 | -0.09 | 0.08 |
| *year\_quit* | 0.25 | 0.17 | -0.20 | -0.01 | 0.21 | -0.11 |
| *systolic\_bp* | -0.04 | -0.05 | -0.09 | -0.07 | 0.03 | 0.00 |
| *diastolic\_bp* | 0.15 | 0.16 | -0.10 | -0.09 | 0.00 | 0.01 |
| *heart\_rate* | 0.16 | 0.16 | -0.16 | -0.12 | 0.03 | 0.04 |
| *respiratory\_rate* | 0.02 | 0.02 | -0.14 | -0.12 | 0.00 | 0.05 |
| *oxygen\_sat* | -0.05 | -0.03 | 0.03 | 0.00 | -0.06 | -0.06 |
| *temp* | -0.04 | -0.05 | -0.05 | -0.03 | 0.03 | 0.09 |
| *wbc* | -0.01 | -0.02 | 0.01 | 0.03 | 0.05 | 0.08 |
| *rbc* | 0.73 | 0.80 | -0.53 | -0.47 | -0.07 | 0.08 |
| *hemoglobin* | 1.00 | 0.96 | 0.12 | 0.25 | 0.37 | -0.34 |
| *hematocrit* | 0.96 | 1.00 | 0.07 | 0.09 | 0.09 | -0.18 |
| *mcv* | 0.12 | 0.07 | 1.00 | 0.92 | 0.21 | -0.37 |
| *mch* | 0.25 | 0.09 | 0.92 | 1.00 | 0.57 | -0.54 |
| *mchc* | 0.37 | 0.09 | 0.21 | 0.57 | 1.00 | -0.60 |
| *rdw* | -0.34 | -0.18 | -0.37 | -0.54 | -0.60 | 1.00 |
| *platelet\_count* | -0.17 | -0.15 | -0.18 | -0.19 | -0.10 | 0.09 |
| *pt* | -0.17 | -0.14 | 0.02 | -0.04 | -0.15 | 0.16 |
| *alt* | 0.08 | 0.05 | -0.04 | 0.01 | 0.11 | -0.08 |
| *ast* | 0.02 | 0.01 | 0.00 | 0.02 | 0.04 | -0.01 |
| *serum\_creatinine* | -0.11 | -0.08 | 0.14 | 0.06 | -0.14 | 0.15 |
| *sodium* | 0.04 | 0.16 | 0.24 | 0.05 | -0.38 | 0.19 |
| *potassium* | -0.08 | -0.02 | 0.12 | 0.01 | -0.21 | 0.05 |
| *total\_serum\_bilirubin* | 0.20 | 0.17 | 0.09 | 0.13 | 0.13 | -0.05 |
| *lactate* | 0.17 | 0.20 | -0.01 | -0.02 | -0.04 | 0.13 |
| *inr* | -0.16 | -0.13 | 0.02 | -0.04 | -0.15 | 0.15 |
| *d\_dimer* | -0.22 | -0.23 | 0.14 | 0.11 | 0.00 | 0.08 |
| *crp* | -0.12 | -0.11 | -0.11 | -0.10 | -0.02 | 0.11 |

| Table 4D | *platelet\_count* | *pt* | *alt* | *ast* | *serum\_creatinine* | *sodium* |
| --- | --- | --- | --- | --- | --- | --- |
| *age* | -0.12 | 0.12 | -0.16 | -0.01 | 0.16 | 0.17 |
| *height* | -0.10 | 0.20 | 0.12 | 0.10 | 0.04 | 0.00 |
| *weight* | 0.04 | 0.03 | 0.00 | -0.03 | -0.01 | -0.06 |
| *year\_quit* | -0.33 | -0.18 | -0.01 | 0.05 | -0.05 | -0.15 |
| *systolic\_bp* | -0.01 | -0.02 | 0.00 | 0.00 | 0.02 | -0.11 |
| *diastolic\_bp* | 0.03 | -0.03 | 0.04 | -0.01 | -0.04 | -0.01 |
| *heart\_rate* | 0.10 | -0.05 | 0.10 | 0.02 | -0.11 | -0.01 |
| *respiratory\_rate* | 0.09 | 0.00 | 0.10 | 0.07 | -0.05 | -0.02 |
| *oxygen\_sat* | -0.05 | -0.06 | -0.01 | -0.08 | 0.04 | 0.07 |
| *temp* | -0.16 | -0.07 | -0.01 | -0.05 | 0.04 | -0.05 |
| *wbc* | 0.36 | -0.02 | 0.06 | 0.02 | 0.03 | 0.02 |
| *rbc* | -0.02 | -0.13 | 0.06 | 0.00 | -0.14 | 0.00 |
| *hemoglobin* | -0.17 | -0.17 | 0.08 | 0.02 | -0.11 | 0.04 |
| *hematocrit* | -0.15 | -0.14 | 0.05 | 0.01 | -0.08 | 0.16 |
| *mcv* | -0.18 | 0.02 | -0.04 | 0.00 | 0.14 | 0.24 |
| *mch* | -0.19 | -0.04 | 0.01 | 0.02 | 0.06 | 0.05 |
| *mchc* | -0.10 | -0.15 | 0.11 | 0.04 | -0.14 | -0.38 |
| *rdw* | 0.09 | 0.16 | -0.08 | -0.01 | 0.15 | 0.19 |
| *platelet\_count* | 1.00 | -0.04 | 0.03 | -0.07 | -0.11 | 0.01 |
| *pt* | -0.04 | 1.00 | 0.49 | 0.72 | 0.04 | -0.02 |
| *alt* | 0.03 | 0.49 | 1.00 | 0.80 | -0.05 | -0.07 |
| *ast* | -0.07 | 0.72 | 0.80 | 1.00 | 0.03 | -0.12 |
| *serum\_creatinine* | -0.11 | 0.04 | -0.05 | 0.03 | 1.00 | 0.05 |
| *sodium* | 0.01 | -0.02 | -0.07 | -0.12 | 0.05 | 1.00 |
| *potassium* | -0.08 | 0.24 | 0.03 | 0.16 | 0.27 | 0.01 |
| *total\_serum\_bilirubin* | -0.11 | 0.16 | 0.14 | 0.18 | 0.03 | -0.01 |
| *lactate* | 0.03 | 0.53 | 0.35 | 0.49 | 0.04 | -0.01 |
| *inr* | -0.04 | 1.00 | 0.52 | 0.75 | 0.03 | -0.04 |
| *d\_dimer* | 0.17 | 0.06 | 0.05 | 0.08 | 0.15 | 0.01 |
| *crp* | 0.21 | 0.11 | 0.10 | 0.16 | 0.02 | -0.07 |

| Table 4E | *potassium* | *total\_serum\_bilirubin* | *lactate* | *inr* | *d\_dimer* | *crp* |
| --- | --- | --- | --- | --- | --- | --- |
| *age* | 0.19 | 0.13 | 0.12 | 0.11 | 0.31 | -0.10 |
| *height* | 0.14 | 0.20 | 0.12 | 0.20 | -0.16 | 0.10 |
| *weight* | 0.00 | 0.05 | 0.04 | 0.02 | -0.22 | 0.27 |
| *year\_quit* | 0.02 | 0.10 | 0.22 | -0.17 | -0.68 | -0.49 |
| *systolic\_bp* | 0.03 | 0.01 | 0.04 | -0.01 | -0.03 | -0.01 |
| *diastolic\_bp* | 0.04 | 0.03 | 0.05 | -0.02 | -0.04 | -0.16 |
| *heart\_rate* | -0.11 | -0.01 | 0.15 | -0.05 | 0.05 | 0.06 |
| *respiratory\_rate* | -0.07 | 0.13 | 0.24 | 0.01 | 0.16 | 0.40 |
| *oxygen\_sat* | 0.07 | -0.11 | -0.31 | -0.04 | -0.11 | -0.29 |
| *temp* | -0.11 | -0.08 | -0.16 | -0.08 | -0.02 | -0.09 |
| *wbc* | -0.01 | 0.18 | 0.25 | -0.03 | 0.17 | 0.44 |
| *rbc* | -0.08 | 0.09 | 0.16 | -0.13 | -0.27 | -0.03 |
| *hemoglobin* | -0.08 | 0.20 | 0.17 | -0.16 | -0.22 | -0.12 |
| *hematocrit* | -0.02 | 0.17 | 0.20 | -0.13 | -0.23 | -0.11 |
| *mcv* | 0.12 | 0.09 | -0.01 | 0.02 | 0.14 | -0.11 |
| *mch* | 0.01 | 0.13 | -0.02 | -0.04 | 0.11 | -0.10 |
| *mchc* | -0.21 | 0.13 | -0.04 | -0.15 | 0.00 | -0.02 |
| *rdw* | 0.05 | -0.05 | 0.13 | 0.15 | 0.08 | 0.11 |
| *platelet\_count* | -0.08 | -0.11 | 0.03 | -0.04 | 0.17 | 0.21 |
| *pt* | 0.24 | 0.16 | 0.53 | 1.00 | 0.06 | 0.11 |
| *alt* | 0.03 | 0.14 | 0.35 | 0.52 | 0.05 | 0.10 |
| *ast* | 0.16 | 0.18 | 0.49 | 0.75 | 0.08 | 0.16 |
| *serum\_creatinine* | 0.27 | 0.03 | 0.04 | 0.03 | 0.15 | 0.02 |
| *sodium* | 0.01 | -0.01 | -0.01 | -0.04 | 0.01 | -0.07 |
| *potassium* | 1.00 | 0.00 | 0.13 | 0.24 | 0.05 | -0.02 |
| *total\_serum\_bilirubin* | 0.00 | 1.00 | 0.16 | 0.15 | 0.23 | 0.11 |
| *lactate* | 0.13 | 0.16 | 1.00 | 0.55 | 0.15 | 0.33 |
| *inr* | 0.24 | 0.15 | 0.55 | 1.00 | 0.06 | 0.09 |
| *d\_dimer* | 0.05 | 0.23 | 0.15 | 0.06 | 1.00 | 0.37 |
| *crp* | -0.02 | 0.11 | 0.33 | 0.09 | 0.37 | 1.00 |

To the right is Figure 1, a heatmap representing the correlation matrix of the data. Note that the labels skip every other row/column of the data, but the values are still present (i.e., the first column/row is labeled age, the third column/row is labeled weight, and the second column/row is unlabeled but represents height.)

Chart, scatter chart

Description automatically generated

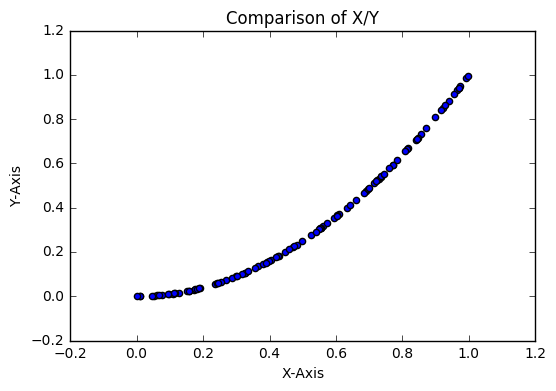
**Figure 0: Heatmap of the correlation matrix.**

1. **DATA SET GRAPHICAL EXPLORATION**

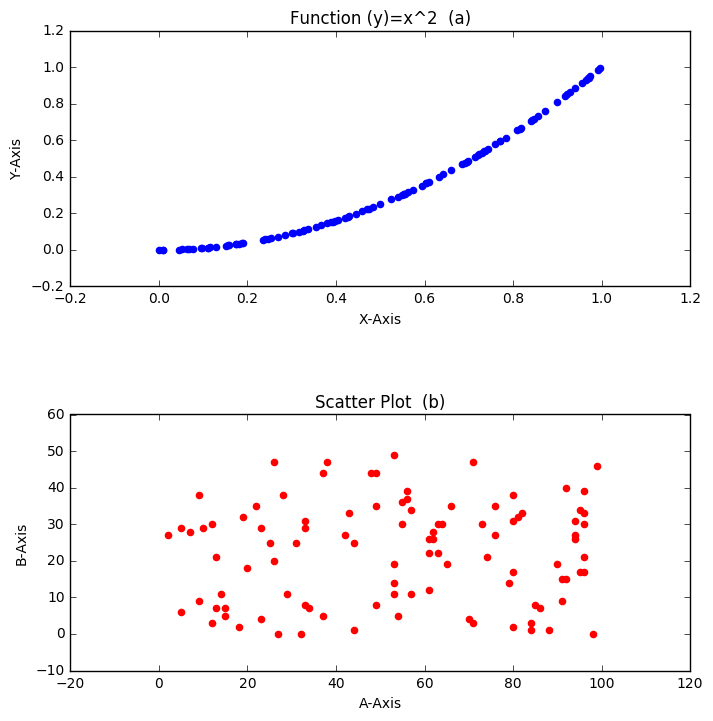
Narrative introduction to the section. In each section below, indicate any interesting distributions, anomalies, imbalance, etc. that you notice.

* 1. *Distributions*
  2. *ScatterPlots / Pairwise Plots (continuous variables)*
  3. *Barcharts (categorical variables)*
  4. *Other Plots - don’t skimp – there are likely other plots that would be useful that I haven’t already specified. Include those in this section.*

All figures should be cited formatted like this and mentioned in the text.



**Figure 1: Comparison of X/Y from dataset (single plot) (8 pt)**



**Figure 2: (a) Function Output (b) A against B (multiple plots) (8 pt)**

1. **SUMMARY OF FINDINGS**

Finish up with a paragraph or two of summarizing your findings about this data set.