**D207 - Increased Hospital Readmission Rates**

(Second Submission)

Jason Smith  
Student ID: 293963  
[jsmi268@wgu.edu](mailto:jsmi268@wgu.edu)  
WGU – MSDA  
D207- Data Exploration of Hospital Data  
Program Mentor: Christiana Okhipo

**Part I:**

**A1. Research question:**

This study will attempt to answer the following question:

Can hospital patient readmissions be predicted based on the services received at the hospital during his or her initial visit? Are there additional indicators (variables) of hospital readmissions that may prove useful?

**A2. Stakeholder Benefit**

Hospital executives are under increased scrutiny when patients have unplanned readmissions within 30 days after the initial visit. The Centers for Medicare and Medicaid Services (CMS), under the Hospitals Readmissions Reductions Program, may withhold regular reimbursements for excessive 30-day readmissions for select diagnoses (Saunders, 2012, Introduction).

Stakeholders can implement steps to reduce hospital readmission rates if they know which patients are most likely to be readmitted and if they know which factors potentially lead to readmissions.

**A3.  Data Identification**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Data Type** | **Description** | **Example** |
| **Customer\_id** | int64 | Unique identification number for each patient | C412403 |
| **Interaction** | object | Unique identification of patient transactions, procedures, and admissions | 8cd49b13-f45a-4b47-a2bd-173ffa932c2f |
| **UID** | object | Unique IDs for tracking transactions, procedures, and admissions | 3a83ddb66e2ae73798bdf1d705dc0932 |
| **City** | object | City of patient’s billing address | Eva |
| **State** | object | State of patient’s billing | AL |
| **County** | object | County of patient’s billing address | Morgan |
| **Zip** | int64 | Zip code patient’s billing address | 35621 |
| **Latitude** | float64 | GPS coordinates of patient billing address | 34.3496 |
| **Longitude** | float64 | GPS coordinates of patient billing address | -86.72508295 |
| **Population** | int64 | Population density within 1 mile of patient’s residence (census data) | 2951 |
| **Area** | object | General area description. Acceptable values: rural, urban, suburban | Suburban |
| **Timezone** | object | Time zone of patient’s | America/Chicago |
| **Job** | object | Employment title of patient or primary insurance policy holder | Psychologist, sport and exercise |
| **Education** | object | Highest level of education received by patient | Some College, Less than 1 Year |
| **Employment** | object | Patient’s employment status. Acceptable values: Full Time, Retired, Student, Unemployed, Part Time | Some College, Less than 1 Year |
| **Marital** | object | Patient’s or primary insurance holder’s marital status. Acceptable values: Divorced, Married, Never Married, Separated, Widowed | Divorced |
| **Gender** | object | How patient describes their gender. Acceptable values: Female, 'Male, Prefer not to answer | Male |
| **ReAdmis** | object | Describes whether the patient was readmitted within the previous month. Acceptable values: yes, no | No |
| **VitD\_levels** | float64 | Measured vitamin D levels of the patient | 17.80233 |
| **Doc\_visits** | int64 | Number of patient visits to the primary care physician | 6 |
| **Full\_meals\_eaten** | int64 | Full meals eaten by patient during hospital stay | 0 |
| **VitD\_supp** | int64 | Number of vitamin supplements per day given to patient | 0 |
| **Soft\_drink** | object | Whether patient consumes 3 or more sodas per day. Acceptable values: yes, no | N0 |
| **Initial\_admin** | object | How was patient admitted to the hospital. Acceptable values: emergency admission, elective admission, observation | Emergency Admission |
| **HighBlood** | object | Patient’s history of high blood pressure. Acceptable values: yes, no | Yes |
| **Stroke** | object | Patient’s history of stroke. Acceptable values: yes, no | No |
| **Complication\_risk** | object | Patient’s level of complication. Acceptable values: high, medium, low | Medium |
| **Arthritis** | object | Patient affected by. Acceptable values: yes, no | Yes |
| **Diabetes** | object | Patient affected by Diabetes. Acceptable values: yes, no | Yes |
| **Hyperlipidemia** | object | Patient affected by Hyperlipidemia. Acceptable values: yes, no | No |
| **BackPain** | object | Patient affected by back pain. Acceptable values: yes, no | No |
| **Allergic\_rhinitis** | object | Patient affected by allergic rhinitis. Acceptable values: yes, no | Yes |
| **Reflux\_esophagitis** | object | Patient affected by reflux esophagitis. Acceptable values: yes, no | No |
| **Asthma** | object | Patient affected by asthma. Acceptable values: yes, no | No |
| **Services** | object | Hospital services received by patient while hospitalized. Acceptable values: blood work, intravenous, CT scan, MRI | Blood Work |
| **TotalCharge** | float64 | Daily hospital charges billed to patient | 3191.048774 |
| **Additional\_charges** | float64 | Additional hospital charges billed to patient | 17939.40342 |
| **Admissions** | int64 | Special charges billed to patient | 3 |
| **Treatment** | int64 | Survey response relating to Timely treatment. Acceptable values: 1-8 | 3 |
| **Visits** | int64 | Survey response relating to Timely visits. Acceptable values: 1-8 | 3 |
| **Reliability** | int64 | Survey response relating to Reliability. Acceptable values: 1-8 | 4 |
| **Options** | int64 | Survey response relating to Options. Acceptable values: 1-8 | 1 |
| **Hours** | int64 | Survey response relating to Hours of treatment. Acceptable values: 1-8 | 8 |
| **Curteous** | int64 | Survey response relating to Courteous staff. Acceptable values: 1-8 | 2 |
| **Dr\_comm** | int64 | Survey response relating to active listening from doctor. Acceptable values: 1-8 | 4 |
| **Children** | float64 | Number of children living in patient’s household. | 2 |
| **Age** | float64 | Patient’s age | 53 |
| **Income** | float64 | Patient’s annual income | 86575.93 |
| **Overweight** | float64 | Is patient overweight? Acceptable values: yes, no | No |
| **Anxiety** | float64 | Patient affected by anxiety disorder. Acceptable values: yes, no | Yes |
| **Initial\_days** | int64 | Number of days patient initially stayed at hospital | 10.58577 |

**B. Data Analysis**

**B1. Chi2\_contingency test analysis –** performed on “ReAdmis” and Services

**B2. Output from calculations:**

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated



**B3. Justification:**

The chi2\_contingency test was selected for its suitability to work with categorical variables. In addition, the ***chi2\_contingency*** test can be used with skewed data so long as the sample size is sufficiently large (MchHugh (Jun, 2013, Section: Introduction). The data set contains 9,999 rows which is sufficiently large for this analysis. See python output (B2).

“ReAdmis” is the dependent variable, a binomial and is necessary to answer the question “Can we predict when a patient is likely to be readmitted? “.

**C. Univariate Analysis on two continuous and two categorical variables**

* Continuous: Income, Age
* Category: Treatment, ReAdmis

**Text

Description automatically generated**

**Graphical user interface, application

Description automatically generated with medium confidence**

**A picture containing graphical user interface

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**C1. Represent Findings Visually**

Techniques sourced from: Tutorialspoint (p. 1).

Chart, bar chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Graphical user interface, application

Description automatically generated

A picture containing box and whisker chart

Description automatically generated

**D. Identification of two continuous and two variables Using Multivariate Analysis**

Multiple techniques sourced: Sani Kamal (2019, Multiple pages).

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**D1. Represent Findings VisuallyGraphical user interface

Description automatically generated**

**Graphical user interface

Description automatically generated**

**Chart, box and whisker chart

Description automatically generated**

**Graphical user interface, chart, box and whisker chart

Description automatically generated**

**Chart, box and whisker chart

Description automatically generated**

**Chart, box and whisker chart

Description automatically generated**

**E.  Summarize the implications of your data analysis by doing the following:**

**1.  Discuss the results of the hypothesis test.**

A contingency table was created of both "Services" (independent variable) and "ReAdmis" (dependent variable) factors. The "chi2\_contingency" hypothesis test was selected because of its ability to work with discrete variables. The resulting p-value is 0.031 and alpha is .05. The Null hypothesis is rejected because the p-value is less than alpha, albeit by a small amount. This means there is a significant difference in the distribution between the “Services” factor for "ReAdmis" values. The contingency table shows patients that received either a “CT Scan” or an “MRI” were more likely to be readmitted into the hospital later.

**2. Discuss the limitations of your data analysis.**

The p-value of 0.031 does meet the minimum threshold. This suggests a possible relationship between the services received and readmissions rates which requires further investigation.

**3. Recommend a course of action based.**

Further analysis is required to determine whether a true relationship exists. Hospital staff recommended to monitor patients that received CT or MRI scans in order to identify possible patterns.

**F.  Provide a Panopto video**

The D207 folder is inaccessible from Panopto. I filed two help desk tickets. First ticket is 27272317, filed on 1/3, and is closed but still unresolved. For this reason, I am providing you a link to my Panopto video.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c4c23638-2544-401c-8fe6-ae15010272c4

**Sources for third-party code:**

1. Tutorialspoint (p. 1). ML - Understanding Data with Visualization  
   <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_understanding_data_with_visualization.htm>
2. Sani Kamal (2019, Multiple pages). Introduction to Multivariate Analysis  
   <https://www.kaggle.com/sanikamal/introduction-to-multivariate-analysis>

**Sources**

1. Saunders ND, Nichols SD, Antiporda MA, Johnson K, Walker K, Nilsson R, Graham L, Old M, Klisovic RB, Penza S, Schmidt CR. (October 15, 2011–April 15, 2012). Examination of Unplanned 30-Day Readmissions to a Comprehensive Cancer Hospital. <https://ascopubs.org/doi/10.1200/JOP.2014.001546>
2. Mary MchHugh (Jun, 2013, Section: Introduction). The Chi-square test of independence  
   [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900058](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900058/)