**Option 1: Capstone Project Hypothesis: U.S. Organization**

Crystal Carter

Colorado State University Global Campus

MIS 542-1 Business Analytics

Dr. Justin Bateh

March 6, 2022

Business Summary

The database institute picked for this project is coming from the National Institute on Drug Abuse (NIDA). NIDA is a component of the National Institute of Health, U.S. Department of Health, and Human Services. NIDA can be tracked back to 1935 as part of the U.S Public health Service (USPHS) hospital. The NIDA is a United States federal government research institute. Their mission is to advance science on the causes and consequences of drug use and addiction and to apply that knowledge to improve individuals and public health. The institute focuses heavily on addictions based on biological, behavioral, and social components. The institute also supports rehabilitation treatments and preventive use of drugs and the consequences. NIDA addresses questions from detecting and responding to emerging drug abuse trends and the understanding of how drugs affect the brain and body. NIDA supports most of the world's research for the health affect drugs and addictions. The institute also supports a large variety of programs to inform policies, improve practice, and advance addition science.

The NIDA partners with different people around the world to find evidence-based solutions to public health problems of drug abuse and additions. The institute has many partnerships with countries, organizations, and individual researchers to support new initiatives research, build international research capacity, and disseminate knowledge. According to National Institute of Healt (NIH) One of NIDA’s most important achievements has been the use of science to clarify central concepts in the field of drug abuse. When NIDA began, correct approaches to drug policy and drug treatment were often thought to hinge on determining whether a particular drug was "physically addicting" or only "psychologically addicting." We now know that addiction has biological, behavioral, and social components. It is best defined as a chronic, relapsing brain disorder characterized by compulsive, often uncontrollable drug craving, seeking, and use, even in the face of negative health and social consequences. NIDA-supported research has also shown that this compulsion results from specific drug effects in the brain. This definition opens the way for broad strategies and common approaches to all drug addiction.

This organization was chosen based on the large scale of how much prescription drugs and street drug abuse are happening all over the United States. According to the 2005 National Survey on Drug Use and Health, the incidence of new nonmedical users of pain relievers is now at 2.2 million Americans aged 12 and older, surpassing the number of new marijuana abusers (2.1 million). Prescription drugs are classified as controlled dangerous substances are widely accepted as essential therapeutic modalities in treating different health issues. However, many prescription medications have appealing side effects that can appeal to one is likes to take these medications outside of the standard prescribed directions by clinicians. The misuse leads to the abuse of prescription drugs can contribute to addictive behaviors serious health risks such as death.

The National Institute on Drug Abuse (NID) and other state and local agencies across the country maintain a close eye for increasing patterns and trends involving prescription drug abuse. A news release stated that NIDA, as a component of the National Institutes of Health, initiated its first large-scale national study related to prescription drug abuse, thereby recognizing it as a serious healthcare issue facing the United States (2007). Nora D. Volkow, MD, Director of NIDA, addressed these concerns at the Congressional Caucus on Prescription Drug Abuse. Nora Volkow's sworn statement declared that nonmedical use of prescription drugs occurs among 7 million Americans per month. The amount surpassed the number of Americans abusing cocaine, heroin, hallucinogens, and inhalants combined (2005).

The data picked for this research project is based on individuals with either a drug abuse or addition problem. The individuals from the data have gone through a period of drug abuse of prescription drugs and have become addicted. The data also exams individuals with street abuse and additions. This data is essential because most of the data research to date has been done on treatment for those abusing street drugs and the individuals addicted to prescription medications. The dataset will provide insight into whether individuals with drug abuse and addictions benefit from outlining and buprenorphine. The increase of abuse and additions in prescription drugs has increased over the years. With the help of NIDA, bringing awareness to the problem will help clinicians and individuals better understand the number of increased health risks there are when it comes to prescriptions and street drug use. The National Overdose Deaths Involving Any Opioid-involved overdose deaths rose from 21,088 in 2010 to 47,600 in 2017 and remained steady in 2018 with 46,802 deaths. This was followed by a significant increase through 2020 to 68,630 overdose deaths.

The dataset is based on individuals working through a trial period for the first steps to stopping drug abuse. The dataset is pulling information from individuals over Most of the data is based on the enrollment of each individualized drug counseling program when they get their prescriptions to compare the effectiveness of different behavioral therapies in conjunction with the medication. The other data have individuals with a brief time of drug counseling. It extremely important to get control of the issues behind the increasing numbers of abuse and additions of prescriptions and street drugs. The National Drug overdose deaths involving prescription opioids rose from 3,442 in 1999 to 17,029 in 2017. From 2017 to 2019, the number of deaths declined to 14,139, followed by an increase to 16,416 in 2020. The numbers of deaths increase two percent from 2019 to 2020.

Business Benefits

Drug abuse and addiction is less about the type or amount of the substance consumed or the frequency of your drug use, and more about the consequences of that drug use. Having research about drug abuse and addictions will help understand the relationship between life issues created by the use and abuse of drugs. The NIDA will benefit from this dataset by better understanding how the effectiveness of buprenorphine/naloxone tablets, marketed as Suboxone, along with different models of drug counseling in patients addicted to prescription opioids. Buprenorphine works by acting on the brain's opiate receptors. It can help target individuals with heroin, morphine, and prescription abuse. Buprenorphine helps relieve an individual's drug cravings without simulating the same intensity of dangerous side effects as the abuse of prescriptions and drug abuse. The combination of naloxone and buprenorphine's abuse potential is limited since those who inject it get high experience severe withdrawal symptoms.

Data Description

Research improves the effectiveness of substance abuse and admissions to hospital admission related to substance abuse. The first insight of the project will be retention in treatment that is associated with better comes, and a principal goal of the dataset is to keep individuals engaged in the treatment and moving towards lower substance abuse. The second insight the dataset will provide the NIDA is treatment success rated based on partnering with clinicians to provide patients with the correct tools and resources to help them overcome drug abuse and addiction. Since drug abuse and addiction are less about the type or amount and more the consequence, it is good to know the difference between drug abuse and addiction individuals. Drug abuse is when an individual uses legal or illegal substance is not standard way. The individual might take the drug more often to feel satisfied with emotional support. However, drug abuse individuals can stop without outside intervention. Drug addiction individuals cannot stop the usage of the drug, regardless of the drug's impact on an individual's health, finance, or emotions. Individuals cannot stop independently. These individuals need help from outside sources.

As of right now, the drug abuse and addiction data are minimal, with a low volume of information. Over the next five weeks, the research looks for better data quality. The current data found from Data World is based on admission to the hospital with drug or alcohol-related issues. The data limitation will impact the outcome, so it will take time to add to the data to ensure we achieve the highest quality of information from data to make future decisions or encourage more research for drug abuse and additions. Given these facts that there are 22 million individuals in the United States are suffering from a substance abuse issue. This number shows the research is out there. We must pull valid data to improve the awareness of how drug abuse and addiction impact the individuals suffering from the disease and the entry family and community where they live. In 2017 there were over 70,000 drug dose-related deaths, and it is estimated that the crime, lost productivity, and health care costs of abuse of drugs alcohol are over 500 billion annually. These are mouth-dropping facts from Psychology Today, and the drug is a massive problem in the United States. Providing insight on these facts and drilling more research on the current problem will ensure that people, families, and the entire United States know how to get help before it becomes a life-altering event.

Data Analytics Tools

Data analytics helps everything from personalized marketing to an individual's life risks. Data analytic tools convert unstructured data into a focused and straightforward way to understand visual representations of critical metrics for research projects, business decisions, risks, and security. These tools enable the raw data to quickly turn into insight that can help understand more about diseases and businesses. In this project, the data used is to prove insight into the number of people suffering from drug abuse and addictions. The data will be prepared to be ready for study and interpretation. Data analytics tools enable aggregate and enterprise data into one viewing since the data will be coming from many sources. The data tools picked for this research project are designed to handle big data from multiple sources. The tools will help interpret data on drug abuse and addiction. It will translate the data into a vivid storytelling visualization to strengthen insight on the issues and create a deeper insight. The tools used for this research project on drug abuse and addiction will be Statistical Analysis System (SAS), Tableau Software, PostgreSQL, and R. The wide variety of data contains valuable insight, and data analytics is the way to unlock research.

These data analytic tools are high-quality tools that will bring together all the data collected throughout the project to ensure a deeper insight into current drug abuse and addictions issues. SAS is a statistical software suite developed by SAS Institute. The tool provides data management, advanced analytics, multivariate analysis, business intelligence, and predictive analytics. SAS helps support the project research manage and alter the data from the different data sources. Tableau Software will provide a visualize the data on drug abuse and addictions. It will convert textual and numerical data to a beautiful interactive dashboard. Tableau provides real-time analysis, collaborations of data, and data blending. The third tool for the project is Postgre SQL. This open-source relational database management system emphasizes SQL compliance. Postgre SQL is designed to handle all ranges of data loads. Postgre SQL is a flexible database with extensive in-depth language support to bring to the project. The last data analytic tool used for this project is the R programming language used for statistical computing. The tool will help build a more substantial database for drug abuse and addictions.

Data Type

In statistical research, a variable is described as an attribute of an object of study. Picking the correct variable type is particularly important to a good experimental design. Knowing the different variables within the data is also essential when choosing the appropriate statistical tool and interpreting the results. The type of data pulled for this project is categorical variables. There are four types of categorical binary, nominal, ordinal variables, and intervals. The category binary is expressed under the bases of a two-numeral system. The usual numbers are zero (0) or one (1). The nominal variables are types of categorical data to represent named qualities. Ordinal variables describe the name qualities of things. Also, they have a natural ranking order amongst them. Intervals data are a type of numerical data. It represents measurements of quantities. Looking over the current data on drug abuse and addictions, the type of data is interval data. As of now, the volume of data is low, so this could change through research for more data on these issues.

Data Dictionary

IBM defines a data dictionary as a centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format (1993). A data dictionary will provide more detail about a research concept, standard definitions of data elements, the meanings, and provides values. The data dictionary is the textual description of data research and its interrelationships. The everyday use of data dictionaries confirms data requirements and for database developers to create and support database systems. It provides the physical attributes of the data elements. Data dictionaries also provide those who build systems and applications that support the data. Lastly, if there is a common, vetted, and documented data resource, it is unnecessary to produce separate documentation for each implementation.

These data's potential will help release valuable information about drug abuse and addictions. The data will help research a more in-depth understanding of the correlation between drug abuse and addictions. In the United States, there is an exceedingly high rate of drug abuse and addictions. They find where individuals compare with adverse events that lead them to start abusing drugs and becoming addicted. Research has already proven that individuals with a mental disorder have a higher chance of abusing drugs. The data shows high drug abuse rates with individuals diagnosed with depression and bipolar disorder, attention-deficit hyperactivity disorder (ADHD), psychotic illness, borderline personality disorder, and antisocial personality disorder (2013). The data can establish causality or directional the cause and effects on an individual's abuse of drugs and addictions.

Future State of The Data

The plans for drug abuse and addictions can provide three paths to insight into mental disorders and drug abuse. The first one would figure out the common risk factors contributing to mental disorders, drug abuse, and addictions. The second mental disorder may contribute to substance use and addiction. Lastly, drug abuse and addiction can contribute to the development of mental disorders. Both drug abuse and mental disorder have been proven to cause overlapping factors such as genetic and epigenetic vulnerabilities, along with similar areas of the brain and the influence of peer pressure with early exposure to mental disorders (2012). Many areas of the brain are affected by drug abuse and addictions. The most significant impact is the brain's circuits, which control decision making, impulse control, and emotional reactions. Many peers' pressures are associated with the increasing risk of drug abuse and addictions, especially with individuals with mental disorders. Stress and pressure of life problems can increase the risk of drug abuse and addictions. The brain responses to stress by mediated through the hypothalamic-pituitary-adrenal (HPA) axis, which in turn can influence brain circuits. This part of the brain controls an individual's motivation. Stress has been shown to reduce activity in the brain, especially the prefrontal cortex, and increase responsivity in the striatum, which leads to decreased behavioral control and increased impulsivity (2008).

Hypothesis Overview

Using a statistical hypothesis for this project assumes the data, which may or may not be accurate. Hypothesis testing allows for formal procedures used by statistical tools to either accept or reject the statistical hypothesis. It is common to use two types of statistical hypotheses. Learn statistics describes statistic hypothesis types with these two methods. One is the null hypothesis, H0 represents a hypothesis of chance basis, and alternative Ha represents a hypothesis of observations influenced by some non-random cause. There is a formal process to use a statistician to determine whether to reject the null hypothesis. In creating the hypothesis testing, following these four steps will help to ensure the consists with the hypothesis of the drugs abuse and addictions. The first step is to state the hypotheses, formulate an analysis plan, analyze sample data, and interpret results. Using these four will clearly understand the data pulled on drug abuse and addiction. Hypothesis testing is significant to obtain the confidence intervals, and every confidence interval can be obtained by a hypothesis test based on the significance of each one

Hypothesis

1. H0 Drug abuse or addiction is not more than five percent of the United States population.
2. Ha Most of the United States population is not suffering from drug abuse or addiction.
3. H0 Drug abuse and addiction is having a ten percent impact on quality of life.
4. Ha Better quality of life in Untied States is more than number of individuals suffering from drug abuse or addiction.

The state hypotheses for drug abuse and addiction are equal to five percent of the population with these current illnesses and are not equal to five percent more. The second hypothesis is that drug abuse and addiction are equal to five percent high but are not greater than population value. Three hypotheses of the data on drug abuse and addiction are not equal to five percent higher and are much higher than population value. This is the best practice to examine the data of a hypothesis test about a population mean. Formulating an analysis plan for drug abuse and addiction data is the value of five percent higher or lower depending on the null hypothesis. Using the correct statistical tool will capture the drug abuse and addiction data. When it comes to analyzing data, there are two types of drug illness: abuse and two addictions. Launching the obtained sample of the different hypothesis means will target the data's drug abuse and addiction error. Before performing any t-testing, the t-test assumptions must be checked. T-test assumes they are the observations of the independent variables that either the observations follow the normal distribution or the same sample size is sufficiently more extensive. These assumptions help for the targeted data error in the experiment. Without doing this step can lead to incorrect conclusions. Interpreting the results of the hypothesis testing helps to provide insight to the data sample and evaluate the claim about the larger population.

A perfect decision guide would always reject the null hypothesis when the current population means is targeted issues errors are less than five percent and always set to fail and reject the null hypothesis when it is not. The drug abuse and addiction data could lead to different samples and conclusions based on what the hypothesis asks for the data to show. One way to ensure the decision is correct would be to reject the null hypothesis that the population means is equal to five percent more which is in favor of the alternative hypothesis that drug abuse and addictions are less than five percent of the data population. One decision that can be concluded from this statement is to reject the claim that drug abuse and addiction data is equal to five percent more of the population. A perfect decision rule does not exist, and some statistics quantify the risks of creating invalid conclusions. Each hypothesis is set up to reject the null hypothesis when the null hypothesis is true, reflecting the worst possible outcome. Setting up the hypothesis test includes two characterizing of type I and type II error rates. The significance level, the probability of type I error, should always be set prior to the data collection.

A few of the commonly used values for running a hypothesis study uses 0.01, 0.05,0.1 and 0.2 by choosing the proper consequences of rejecting the null hypothesis when true. Type I error means concluding the target location error is less than five percent when it is five percent. Type II errors can depend on the delta, the significant operational difference between the two. During the testing period, the statistical tools will help to determine the difference in the mean targets of drug abuse and addiction data. We will use histograms, linear regression models, and binary linear regression to run these values. The histogram model will be used during drug abuse and addictions research. Histogram model charts the distribution of a numeric variable's values in a series of bars. The bars cover the range of numeric values, called bins of classes. The height of the bars indicates the frequency of data points within a value within the corresponding bin. Histograms are good to show general distributional features of the drug abuse and addiction data. It demonstrates the difference between values regardless of their absolute values.

Using a linear regression model is one way to analyze the drug abuse and addiction data by considering the linear relationship between the dependent variables and one or more independent variables. The linear regression is used to show the strength of the relationship of correlation between numerous factors and the dispersion of the results to explain the behavior of the dependent variables. In this case, the dependent variables are drug abuse and addiction. The independent can be the different drugs, mental illnesses, and age range. The linear regression model's goal will be to estimate the magnitude of a relationship between variables and whether it is statistically significant. Once the linear regression model runs, the model will populate an output. Another goal of the regression model is to obtain a line equation that best fits the data. The best fit line is the total prediction error for all the data points, which is considered as small as possible. The error is the distance between the points on the plane to the best fit line.

To interpret the model, a few things to understand produce different numerical results. The coefficients tell us the association between independent and dependent variables. For example, if the coefficient is +0.12, it tells us that every 1-point change in the variable corresponds with a 0.12 change in the independent variable in the same direction. If the output were –3.00, it would tell us that a 1-point change in the explanatory variable results in a 3x change in the dependent variable, in the opposite direction. To determine if the model is statistically significant will be based on each coefficient's standard error, such as the p-value and confidence intervals. When analysts use the p-value of 0.05 or less, it indicates significance, and if the p-value is more significant, we cannot rule out the chance for the coefficient results.

Reference

Compton, W., Thomas, Y., Stinson, F., & Grant, B. (2013). Prevalence in the United States Results from the National Epidemiologic Survey on Drug and Related Conditions. Arch Gen Psychiatry.

Estevez, E. (2022 February 21). Creating a Linear Regression Model in Excel. Investopedia. Retrieved on March 7, 2022, from [Linear Regression Excel: Step-by-Step Instructions (investopedia.com)](https://www.investopedia.com/ask/answers/062215/how-can-i-run-linear-and-multiple-regressions-excel.asp#:~:text=Linear%20regression%20models%20the%20relationship%20between%20a%20dependent,best%20fit%20among%20all%20variables%20in%20the%20model.)

Hagan, E. (2018 September 26). The Drug Problem in the U.S. Is Not What We Think It Is. Psychology Today. Retrieved on February 25, 2022 from [The Drug Problem in the U.S. Is Not What We Think](https://www.psychologytoday.com/us/blog/crime-and-punishment/201809/the-drug-problem-in-the-us-is-not-what-we-think-it-is#:~:text=At%20any%20given%20moment%2C%20there%20are%20about%204.5,deaths%20in%202017%2C%20the%20majority%20due%20to%20opioids.)  It Is | Psychology Today

Important Events in NIDA History. NIH. Gov. Retrieved on February 21, 2022, from [National Institute on](https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-institute-drug-abuse-nida#programs)  Drug Abuse (NIDA) | National Institutes of Health (NIH)

Mcgraw. (1993). IBM Dictionary of Computing p758. Professional Book Group. [Statistics - Hypothesis](https://www.tutorialspoint.com/statistics/hypothesis_testing.htm#:~:text=%20Statistical%20hypotheses%20are%20of%20two%20types%3A%20,are%20influenced%20by%20some%20non-random%20cause.%20More%20)  testing (tutorialspoint.com)

Pelayo-Teran, J., Suarez-Pinilla, P., Chadi, N., & Crespo-Facorro, B. (2012). Gene-Environment Interactions Underlying the Effect of Cannabis in Frist Episode Psychosis. *Cur Pharm Des.*

(2007 November 3) National Insititute on Drugs Abuse. NIDA Launches First Large-Scale National Study to Treat Addiction to Prescription Pain Medications. Retrieved on Feb 21, 2022 from <https://www.archives.drugsabuse.gov/newsroom/07/NR3-07/html>

(2005). National Institute on Drug Abuse. Congressional Caucus on Prescription Drug Abuse. Retrieved on Feb 21, 2022 from <https://ada.org/sections/about/pdfs/ada.pdf>

National Institute on Drug Abuse (NIDA). The NIH Almanac. National Institutes of Health. Retrieved on February 20, 2022, from [National Institute on Drug Abuse (NIDA) | National Institutes of Health](https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-institute-drug-abuse-nida#:~:text=Mission%20The%20mission%20of%20the%20National%20Institute%20on,that%20knowledge%20to%20improve%20individual%20and%20public%20health.)  (NIH)

Overdose Death Rates. National Institute on Drug Abuse. Retrieved on February 22, 2022, from [Overdose Death Rates | National Institute on Drug Abuse (NIDA) (nih.gov)](https://nida.nih.gov/drug-topics/trends-statistics/overdose-death-rates)

Renze, M. (2019 June 15). Nominal, Ordinal, Interval, and Ratio Data. Retrieved on February 25, 2022, From [Nominal, Ordinal, Interval, and Ratio Data - Matthewrenze](https://matthewrenze.com/articles/the-four-subtypes-of-data-in-data-science/#:~:text=Nominal%2C%20Ordinal%2C%20Interval%2C%20and%20Ratio%20Data%201%20Nominal,data%20are%20a%20type%20of%20numerical%20data.%20)

Sinha, R. (2008). Chronic Stress, Drug Use and Vulnerability to Addiction. *Ann N.Y. Acad Sci*

Statistics Hypothesis Testing. Learn Statistics. Tutorials Point. Retrieved on March 7, 2022, from [Statistics](https://www.tutorialspoint.com/statistics/hypothesis_testing.htm)  - Hypothesis testing (tutorialspoint.com)