

NicoTeen Wars

1 Executive Summary

Not many things short of hypnotism can impel a person to surrender their home, career, personal relationships, and future. Illicit substances not only have the ability to ruin a person's life, but the ongoing search for the next high can effectively destroy entire families and communities. The vehement pull of narcotics is far too easy to give in to under today's societal conditions. Illegal substances are increasingly available to youth and have begun to be considered a staple of American teen life.

With the influence of illicit substances taking global population by storm, nations and governments must consider how to properly combat an issue has not only financial implications such as health care, the criminal justice system, and workplace productivity, but also non-financial ramifications such as divorce and domestic abuse. The specific marketing strategies of e-cigarette companies that target the youth have prompted increased focus and governmental restrictions intended to protect underage children from the pervasiveness of e-cigarettes in modern culture. Despite this, e-cigarette usage in children has increased consistently across all high school ages over the course of 2011 to 2015. The development of quantitative models to analyze the applicability and reliability are imperative to assuring that redistribution strategies can succeed in meeting the needs of the food-insecure. In order to meet this need, we have developed a variety of models which aim to vape usage and the spread of nicotine usage over the next four years, identifying the primary social influences and characteristics of the individual and the drug itself, and developing a robust metric for the impact of substance use to quantify and order the degree of severity of drugs: nicotine, marijuana, alcohol, and un-prescribed opioids.

First, we developed a model that predicts the use of nicotine products due to vaping over the next 10 years. This model is based on the rise of nicotine products due to cigarettes from the 1900s. It was created in Python and extrapolates data from the current rise of vape products beginning in the 2000s.

We then used data on the proportions of individuals in America by gender, race/ethnicity, and income distribution and the probability of an individual given the aforementioned characteristics of using and abusing alcohol, nicotine, opioids, and marijuana to create a multivariate model to predict the likely number of individuals in a class of 300 to use these substances. This model was created in Python, and it can be used to identify the key demographics in which substance abuse is most prevalent.

Finally, we developed a set of criteria that we decided would be factors to affect the likelihood of a person using a particular substance. The factors that we took into consideration were loss of tangible goods, economic cost, mortality based on specific drug use, likelihood of addiction, effect on others, physical harm, incarceration rate, and average social index. We then applied a prevalence multiplier and prevalence multiplier on social impacts, as well as the average individual index, all in order to adequately adjust the scores of each substance. After researching these factors in relation to the specific substances of alcohol, heroin (as representation of opiates), nicotine, and marijuana, we scored each substance on each category out of 100, then added and adjusted in order to come up with a final score metric, which we deemed a MEJJJ Index, to finally rank each substance based on the values. The higher the MEJJJ Index, the higher the impact of the specific substance.

Overall, our investigation revealed that efforts must be taken to combat the increasing demand and usage of vape products among today's youth, that white males were determined to be most likely to use the substances in question, although there was variation among demographics across the four different substance, and that alcohol is currently the most significant drug American culture followed by heroin, tobacco and marijuana.

Contents

1	Executive Summary	1
2	Episode I: The Vapor Menace	3
2.1	Objective	3
2.2	Predictions	3
2.3	Justifications	3
3	Episode II: The Attack of the Drugs	4
3.1	Objective	4
3.2	Local Assumptions	4
3.3	Preliminary Formulas	4
3.4	Solution	4
3.5	Justifications	4
3.6	Data Interpretation	5
4	Episode III: Revenge of the Ranks	6
4.1	Objective	6
4.2	Local Assumptions	6
4.2.1	Data Table	6
4.3	Solution	6
4.4	Justifications	7
5	Appendix	10
5.1	spread.py: Calculating the spread of Vape Usage	10
5.2	drug_analysis.py: Calculating the probability of using a substance	10
5.3	all_students.py: Calculating expected numbers in a class of 300	11
5.4	Alcohol.dat	11
5.5	Marijuana.dat	12
5.6	Opioids.dat	12
5.7	Nicotine.dat	13

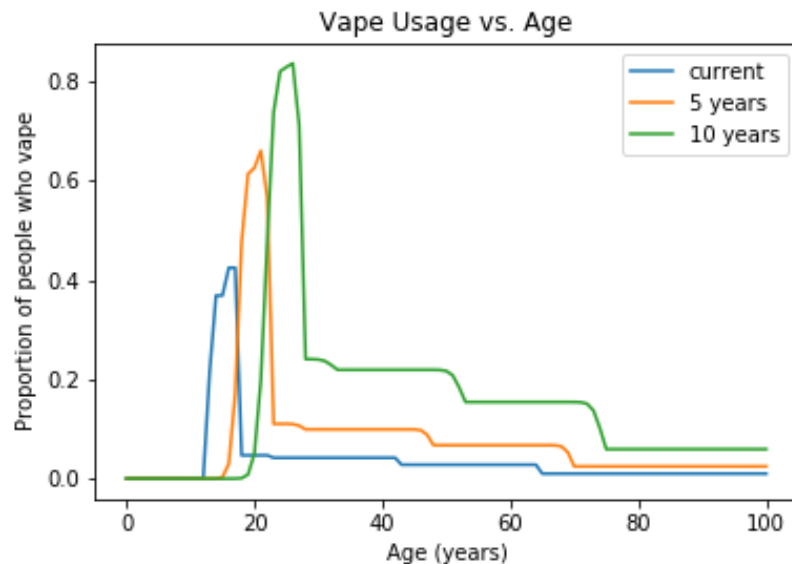
2 Episode I: The Vapor Menace

2.1 Objective

We are asked to create a model that predicts the spread of nicotine use due to vaping over the next decade and analyze how this new form of nicotine compares to the use of cigarettes.

2.2 Predictions

This graph depicts the age distribution of vape usage, currently, 5 years from now, and 10 years from now.



2.3 Justifications

The change in distribution was calculated using a logistic-like formula. Since the pressure to begin to vape is affected by both people older than you as well as by people your own age, the number of people who begin vaping in a given year can be modeled as

$$k * (P(t) + P(t + 1))(1 - P(t))$$

where $P(t)$ is the proportion of people who vape at age t , and k is some constant[4]. Note that it is proportional to $1 - P(t)$ because the the pressure creates a probability to start, thereby causing more people to start if fewer have already. To create the graph above, $k = 0.1$ was used. This thereby models how peer pressure will increase vape usage throughout young age groups, while continuing to increase it among groups that currently vape greatly. This model was calculated using the program in 6.1, the data for which was derived from the data set provided.

This prediction ignores the effect of legislation on vaping, and thus may be overestimating the spread of vaping[2]. As seen with the 1964 Surgeon General's Report on Smoking and Health[19], cigarette use subsequently over time due to new publicly available information on the harmful effects of vaping on health. It can be assumed that the FDA's recent activity on regulating the use of e-cigarettes such as restricting the sale of flavored versions that appear to target children will have a similar effect[32]. There may also be a development of a different recreational product that will become widely popular and replace widespread vaping, just as vapes generally replaced the popularity of cigarettes. Although these factors may cause the use of vapes may decrease due to these health concerns, the current social climate and generation of normalized vaping allows the predictions of an increase in use continue to be relevant.

3 Episode II: The Attack of the Drugs

3.1 Objective

We are asked to create a model that simulates the likelihood of that a given individual will use a given substance.

1. Predict how many students among a high school class of 300 high school seniors with varying characteristics will use the following substances.

- Nicotine
- Marijuana
- Alcohol
- Unprescribed Opioids

3.2 Local Assumptions

1. The ratios of given people who are likely to use a given substances are independent of age.
2. The overdose rate of people are independent of any factors.
3. We assume there to be only two genders, male and female.

3.3 Preliminary Formulas

- For events A and B , $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- For events A , B , and C , $P(A \cap B \cap C) \approx \frac{P(A \cap B)P(A \cap C)}{P(A)}$

3.4 Solution

Number of Alcohol users: 89.54065428636922

Number of Nicotine users: 90.94330121091657

Number of Marijuana users: 65.82127566746348

Number of Opioids users: 1.1381565034848171

These numbers were calculated using the calculations described below, using the code in 6.3. The actual mechanics for calculating the probability given certain conditions is using the code in 6.2 as well as the .dat files in 6.4-6.7. These are calculated by creating a distribution for the 300 students (based on the US averages) and then applying the formulas described. These match almost perfectly with the actual national average of usage among high school seniors. Using the expected values

3.5 Justifications

Any general substance can be said to have factors F_1, F_2, \dots, F_n , which are each variables that take on values for any given individual. For example, any substance can be associated with the factors age, race, and income, which may take on a variety of values. Assumption (1) allows us to use general data on consumption of a particular substance to calculate the proportion of all seniors in high school who use the substance, by scaling by a factor of $\frac{P(\text{Use drug} \mid \text{Senior in HS})}{P(\text{Use drug})}$. Then we can rearrange formulas (1) and (2) in the following manner, where using a particular drug is the event D , and the values for the n factors are V_1, \dots, V_n :

$$P(D|V_1 \cap V_2 \cap \dots \cap V_n) = \frac{P(D \cap V_1 \cap V_2 \cap \dots \cap V_n)}{P(V_1 \cap V_2 \cap \dots \cap V_n)}$$

Assuming that the factors are approximately independent and repeatedly applying (1) to the numerator yields

$$P(D|V_1 \cap V_2 \cap \dots \cap V_n) = \frac{P(D \cap V_1) \dots P(D \cap V_n)}{P(D)^{n-1} P(V_1) \dots P(V_n)}$$

Simplifying all pairs of $\frac{P(D \cap V_i)}{P(V_i)} = P(D|V_i)$ yields

$$P(D|V_1 \cap V_2 \cap \dots \cap V_n) = \frac{1}{P(D)^{n-1}} \prod_{i=1}^n P(D|V_i)$$

However, all of the values for $P(D|V_i)$ do not account for the age, so applying assumption (1) and using the scaling factor mentioned above yields the final formula,

$$P(D|V_1 \cap V_2 \cap \dots \cap V_n) = \frac{P(D|\text{Senior})}{P(D)^{n-1}} \prod_{i=1}^n P(D|V_i)$$

3.6 Data Interpretation

The data for the factors of race, gender, and income for each substance were researched and interpreted to be read by the `drug_analysis.py` program by formatting them into a `.dat` file. In each `.dat` file, the leading value on the first line is the proportion of high school seniors that have used the substance in the past month, and the following lines are groups of 3, where the first line is the category. The line after the category title includes the data subcategories and each value next to the subcategory is the proportion of people in the United States that fit into that category. The next line contains the same subcategories but instead contains proportions of those subcategories that use the substance. These three lines are repeated for different categories. The python program was created to account for varying numbers of categories and subcategories so that the entered data is flexible.

4 Episode III: Revenge of the Ranks

4.1 Objective

We are asked to develop a robust metric for the impact of substance use taking both financial and non-financial factors and rank the substances mentioned in the previous section.

4.2 Local Assumptions

1. Nicotine was considered to be all tobacco products
2. Opioids were considered to be heroine
3. Data was taken from the a study of the United Kingdom and was applied to the United States

4.2.1 Data Table

Data Table 1: Evaluated Statistics				
	Alcohol	Heroin	Nicotine	Marijuana
Loss of Tangibles	50	29	10	12
Economic Cost	17	5	6	4
Drug Specific Mortality	72	55	26	20
Addiction	64	100	74	50
Effect on Others	74	85	47	50
Physical Harm	47	93	41	33
Crime	34	20	8	7
Average Social Index	41.66666667	36.66666667	20.33333333	20.33333333
Prevalence Multiplier	0.864	0.016	0.451	0.37
Prevalence Multiplier On Societal Impacts	36	0.586666667	9.170333333	7.523333333
Average Individual Index	58.25	69.25	37.75	28.75
Final MEJJJ Index	94.25	69.83666667	46.92033333	36.27333333

4.3 Solution

The average of the individual indexes of those subcategories classified as social impacts—Effect on Others, Crime, and Economic Cost—were multiplied by a prevalence multiplier calculated by examining the ubiquity of the four drugs throughout the American population. This approach is modeled by the equation

$$\frac{P_m(E_{oO} + C + E_c)}{3}$$

The average of the individual indexes of those subcategories classified as individual impacts—Physical Harm, Addiction, Morality, and Crime. These values were not multiplied by a prevalence multiplier as these factors are density-independent and are not affected by the proportion of Americans utilizing the given drugs. This approach to assessing the severity of individual issues is modeled by the equation

$$\frac{\text{Physical Harm} + \text{Addiction} + \text{Mortality} + \text{Tangibles}}{4}$$

The Average Individual Index was then added to the Average Social Index to yield a Final MEJJJ Index. By placing the four drugs in numerical order with the highest index corresponding with the metric for the greatest impact of the substances and yields an order of Alcohol, Opioids, Nicotine, and Marijuana.

4.4 Justifications

The calculated standings of our Final MEJJJ Index reveals that, on a holistic view of American society, Alcohol is the most prevalent issue regarding the overall population. The combined individual and societal effects of Alcohol outweigh the effects of Opioids, Nicotine, or Marijuana, implying that America's "deadliest" and most harmful drug is, unsurprisingly, alcohol. The relatively high MEJJJ values for Alcohol and Opioid are valid when considering the realities American culture, which make alcohol and opioids readily available and accessible for potential users at young ages. One strength of the MEJJJ model is the accuracy of the model in relation to American society, where 60% of underage high school teenagers have had an alcoholic beverage in their lifetime [38]. Our model incorporates both holistic and specified analyses for each drug's areas of influence. Another strength of our model is the comprehensible and simplistic use of the model, which allows for users to simply input the values of another drug to receive the desired drug's MEJJJ value. One limitation of our model is the inability for the model to factor in for marginal analysis. The model does not account for changes in population and thus does not make accommodations for changes in the prevalence for a given drug. Consequently, our model's primary use is not to calculate the real impact of substance use using a base year, but is instead useful to assess the impact of substance use for an individual year.

References

- [1] <https://aspe.hhs.gov/system/files/pdf/259261/ASPEEconomicOpportunityOpioidCrisis.pdf>
- [2] <https://www.cdc.gov/tobacco/campaign/tips/resources/data/cigarette-smoking-in-united-states.html>
- [3] https://www.cdc.gov/tobacco/data_statistics/sgr/history/index.htm
- [4] <https://www.cdc.gov/vitalsigns/heroin/infographic.html>
- [5] https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html
- [6] <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6452a3.htm>
- [7] <https://www.cnn.com/2018/10/22/teen-cigarette-smoking-ticks-up-as-vaping-surges.html>
- [8] <https://news.gallup.com/poll/1582/Alcohol-Drinking.aspx>
- [9] <https://news.gallup.com/poll/184358/drinking-highest-among-educated-upper-income-americans.aspx>
- [10] https://news.gallup.com/poll/237839/americans-say-marijuana-vaping-less-harmful-tobacco.aspx?g_source=link_news9g_campaign=item_237818g_medium=copy
- [11] <https://news.gallup.com/poll/214250/say-tried-marijuana.aspx>
- [12] <https://news.gallup.com/poll/237818/young-people-adopt-vaping-smoking-rate-plummets.aspx>
- [13] <https://www.infoplease.com/science-health/health/smoking-prevalence-among-us-adults-1955a2013>
- [14] <https://jamanetwork.com/journals/jama/fullarticle/2428954?resultClick=3>
- [15] <http://monitoringthefuture.org/pubs/monographs/mtf-overview2015.pdf>
- [16] <https://www.statista.com/statistics/261576/cigarette-consumption-per-adult-in-the-us/>
- [17] <https://www.statista.com/topics/1600/smoking/>
- [18] <https://www.narconon.org/blog/drug-use/does-race-gender-or-ethnicity-determine-drug-use/>
- [19] <https://www.nap.edu/read/11795/chapter/4#42>
- [20] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4644488/>
- [21] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3164783/>
- [22] <https://www.ncbi.nlm.nih.gov/books/NBK294302/table/ch13.t2/?report=objectonly>
- [23] <https://www.ncbi.nlm.nih.gov/books/NBK294302/table/ch13.t2/?report=objectonly>
- [24] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3185179/>
- [25] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4872616/>
- [26] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3410945/>
- [27] <https://www.npr.org/sections/health-shots/2015/03/17/393554628/income-affects-how-genes-play-a-role-in-drinking-problems>
- [28] <https://www.samhsa.gov/data/nsduh/reports-detailed-tables-2017-NSDUH>
- [29] <https://vizhub.healthdata.org/tobacco/>
- [30] <http://www.ias.org.uk/uploads/pdf/news%20stories/dnutt-lancet-011110.pdf>
- [31] https://www.researchgate.net/publication/6424313_Development_rational_scale_to_assess_the_harm_of_drugs_of_potential

-
- [32] <https://www.forbes.com/sites/tomangell/2017/09/25/trump-administration-makes-it-harder-to-track-marijuana-arrests-but-i-did-it-anyway/5295f9e468bc>
 - [33] <https://www.alcoholrehabguide.org/alcohol/crimes/>
 - [34] https://www.drugwarfacts.org/chapter/crime_arrests
 - [35] <https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-facts-and-statistics>
 - [36] <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2612444>
 - [37] <https://www.omicsonline.org/lifetime-and-current-prevalence-of-tobacco-smoking-2155-6105.1000145.php?aid=14160>
 - [38] <https://pubs.niaaa.nih.gov/publications/UnderageDrinking/UnderageFact.htm>
 - [39] <https://mym3challenge.siam.org/the-problem>

5 Appendix

5.1 spread.py: Calculating the spread of Vape Usage

```
import pylab
initial = [0] * 13 + [0.215] + [0.369] * 2 + [0.425] * 2 + [0.047] * 5 + [0.042] * 20 + [0.028] * 22 + [0.01] * 36
print(len(initial))
def evolve(arr, k):
    new = [0] * 101
    for i in range(1, len(arr)):
        new[i] = initial[i - 1] + k * (initial[i - 1] + initial[i]) * (1 - initial[i - 1])
    return new

pylab.plot(list(range(len(initial))), initial, label="current")
for i in range(5):
    initial = evolve(initial, 0.1)

pylab.plot(list(range(len(initial))), initial, label="5 years")
for i in range(5):
    initial = evolve(initial, 0.1)

pylab.plot(list(range(len(initial))), initial, label="10 years")
pylab.title("Vape Usage vs. Age")
pylab.xlabel("Age (years)")
pylab.ylabel("Proportion of people who vape")
pylab.legend(loc="upper right")
```

5.2 drug_analysis.py: Calculating the probability of using a substance

```
drug = input("Enter drug name: ")

def process(line1, line2, line3):
    attribute = line1
    att_1 = dict()
    att_2 = dict()
    l2 = line2.split(" ")
    l3 = line3.split(" ")
    for i in range(0, len(l2), 2):
        att_1[l2[i]] = float(l2[i + 1])
        att_2[l3[i]] = float(l3[i + 1])
    print("Enter " + attribute + ":")
    choice = input("Options are " + str([k for k in att_1]) + "\n")
    new_prop = sum([att_1[k] * att_2[k] for k in att_1.keys()])
    specific_prop = att_2[choice]
    return new_prop, specific_prop

try:
    with open((drug + ".dat"), "r") as file:
        lines = file.read().splitlines()
        total_proportion = float(lines[0])
        count = 1
        all_props = 0
        num_variables_counted = 0
        while count + 2 <= len(lines):
            new_prop, specific_prop = process(lines[count], lines[count + 1], lines[count + 2])
            all_props = ((all_props * num_variables_counted) + new_prop) / (num_variables_counted + 1)
            count += 3
            num_variables_counted += 1
            total_proportion *= specific_prop
        total_proportion /= (all_props ** num_variables_counted)
        print(total_proportion)
except:
    print("Error: File not Found")
```

5.3 all_students.py: Calculating expected numbers in a class of 300

```

for drug in ["Alcohol", "Nicotine", "Marijuana", "Opioids"]:
    with open(drug + ".dat", "r") as file:
        lines = file.read().splitlines()
        count = 1
        props1 = []
        props2 = []
        while count + 2 <= len(lines):
            l1 = lines[count + 1].split(" ")
            props1.append([float(l1[i]) for i in range(1, len(l1), 2)])
            l2 = lines[count + 2].split(" ")
            props2.append([float(l2[i]) for i in range(1, len(l2), 2)])
            count += 3
        prop_sums = [sum([props1[i][j] * props2[i][j] for j in range(len(props1[i]))]) for i in range(len(props1))]
        prop = sum(prop_sums)/len(prop_sums)
        # print(prop)
        prop_using = 0
        for v1 in range(len(props1[0])):
            for v2 in range(len(props1[1])):
                for v3 in range(len(props1[2])):
                    prop_using += props2[0][v1]*props2[1][v2]*props2[2][v3]*props1[0][v1]*props1[1][v2]*props1[2][v3]
        #print(prop_using)
        # print(props2[0][0]*props2[1][2]*props2[1][0]/(prop**3)*float(lines[0]))
        prop_using *= float(lines[0])
        prop_using /= prop ** 3
        print("Number of " + drug + " users: " + str(300*prop_using))

```

5.4 Alcohol.dat

0.302

Race

White 0.607 Asian 0.058 Black 0.134 Native_American 0.015 Hispanic 0.181

White 0.598 Asian 0.38 Black 0.438 Native_American 0.478 Hispanic 0.463

Income

0-29999 0.53 30000-74999 0.33 75000- 0.14

0-29999 0.45 30000-74999 0.67 75000- 0.78

Gender

Male 0.492 Female 0.508

Male 0.574 Female 0.65

5.5 Marijuana.dat

```

0.222
Race
White 0.607 Asian 0.058 Black 0.134 Native_American 0.015 Hispanic 0.181
White 0.087 Asian 0.038 Black 0.10 Native_American 0.134 Hispanic 0.084
Income
0-29999 0.53 30000-74999 0.33 75000- 0.14
0-29999 0.42 30000-74999 0.42 75000- 0.44
Gender
Male 0.492 Female 0.508
Male 0.48 Female 0.35

```

5.6 Opioids.dat

```

0.004
Race
White 0.607 Other .393
White 0.03 Other 0.017
Income
0-19999 0.39 20000-49999 0.34 50000- 0.27
0-19999 0.055 20000-49999 0.023 50000- 0.016
Gender
Male 0.492 Female 0.508
Male 0.036 Female 0.016

```

5.7 Nicotine.dat

0.32096

Race

White 0.607 Asian 0.058 Black 0.134 Native_American 0.015 Hispanic 0.181

White 0.166 Asian 0.90 Black 0.165 Native_American 0.318 Hispanic 0.107

Income

0-19999 0.39 20000-49999 0.34 50000-99999 0.19 100000- 0.08

0-19999 0.382 20000-49999 0.296 50000-99999 0.215 100000- 0.173

Gender

Male 0.492 Female 0.508

Male 0.175 Female 0.135