

Trends in Drug Consumption Analysis



ISDS 415
Tue 7:00 to 9:45 pm
(Dr. Rahul Bhaskar)

By:
Juhi Dalal
Kedar Kulkarni
Aniket Shete
Gabby Sierra
Raquel Ramirez
Zach Settler



Data Analysis 1: How do the values of the N-Score rating results compare by cannabis consumption and age?

To begin with, we have made a comparison between three different factors from our data set, these factors are: N-Score, cannabis consumption, and age. The N-score is neuroticism, which is one of the Big-5 personality traits in the study of psychology. Individuals who score high on the N-score are more likely than average to be moody and experience feelings of anxiety, worry, fear, anger, frustration, envy, guilt, depression, and loneliness.

We decided to test these three variables out because we predicted there to be a high correlation between a high rating on N-score and drug consumption, and decided to tie in age as well to give a clearer picture of the data.

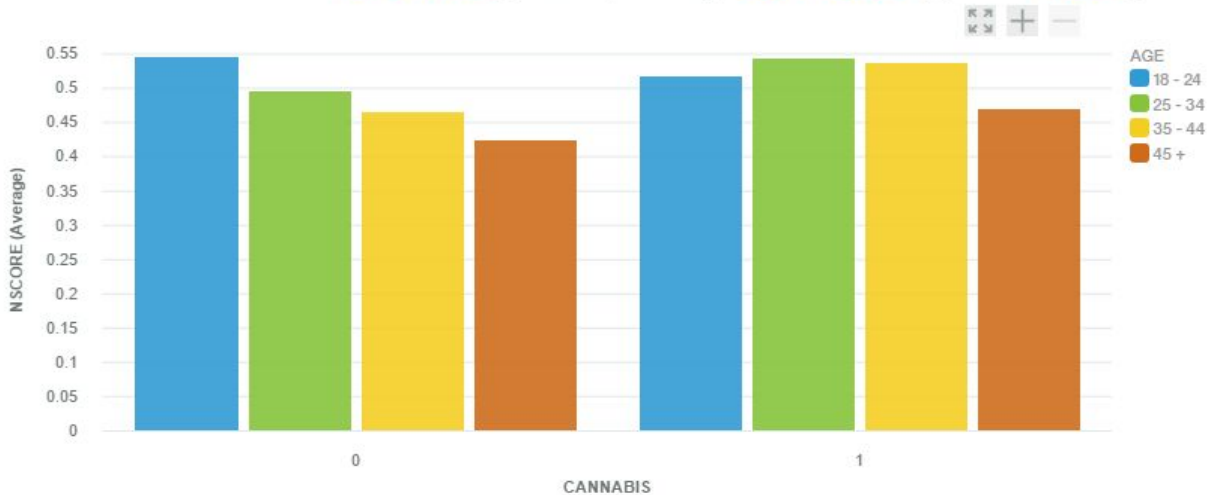
The N-score rating is based on an average rating from 0.00 to 1.00. In addition, cannabis is classified into a "0" being a non-active consumer, and "1" being an active consumer. Lastly, age is split into 4 different categories: 18-24, 24-34, 34-44, and 45+.

As we explored different types of graphs on Watson Analytics, we concluded that a bar graph was the best representation of the data. It is shown below.

The bar graph for inactive cannabis users shows that the average N-score decreases as one gets older. On the other hand, the active user bar graph shows more fluctuation. Those in the 25-34 age group have the highest average N-score, followed by 35-44.

Across both bar graphs, active cannabis users are less neurotic than inactive users in the 18-24 age range, but more neurotic in the 25-34, 35-44, and 45+ age ranges. Because the use of marijuana is becoming legal in several states, this type of data may be useful for marijuana producers who wish to market their products to a certain age group and personality type.

How do the values of **NSCORE** compare by **CANNABIS** and **AGE** ?



Data Analysis 2: How do the values of IMPULSIVITY compare by AGE and COKE?

Impulsivity (or impulsiveness) is a multifactorial construct that involves a tendency to act on a whim, displaying behavior characterized by little or no forethought, reflection, or consideration of the consequences.

People are inclined to act on impulse rather than thought and are motivated by or resulting from impulse.

We did the analysis on how impulsivity is affected if a person is doing drugs. We selected Coke as the drug to see its effect on the impulsivity. We surveyed a thousand people and categorized them by their age.

0 indicates active users of coke and 1 indicates inactive users of coke.

We formed categories for age as follows:

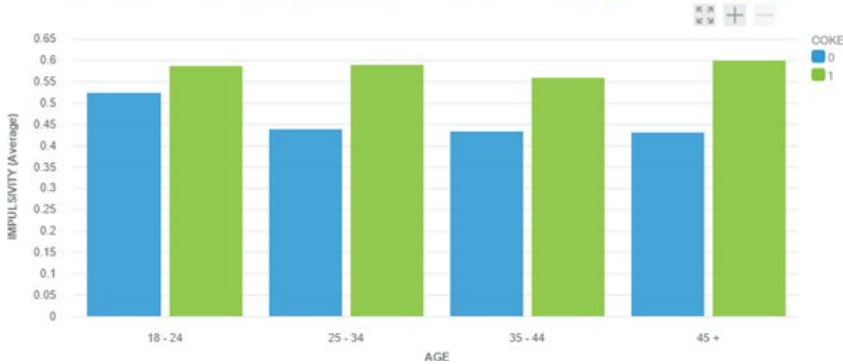
Group 1: From age 18-24, Group 2: From age 25-34

Group 3: From age 35-44, Group 4: From age 45+

Using the available data, we asked a question in Watson Analytics.

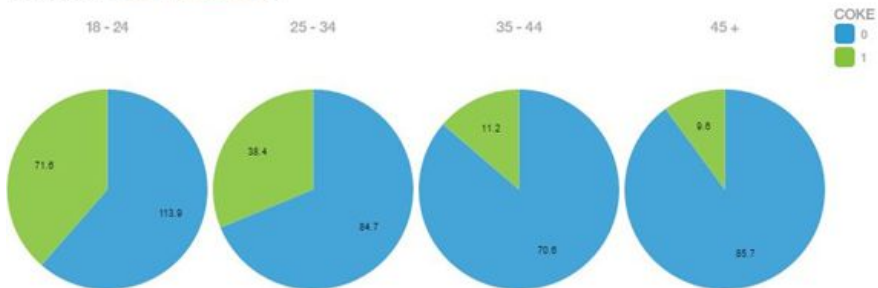
“How do the values of IMPULSIVITY compare by AGE and COKE?” and after considering various types of graphs on Watson Analytics, we selected a bar graph and a pie chart to represent our data.

How do the values of **IMPULSIVITY** compare by **AGE** and **COKE** ?



What is the breakdown of **IMPULSIVITY** by **COKE** ?

Filtered by AGE: 4 selected



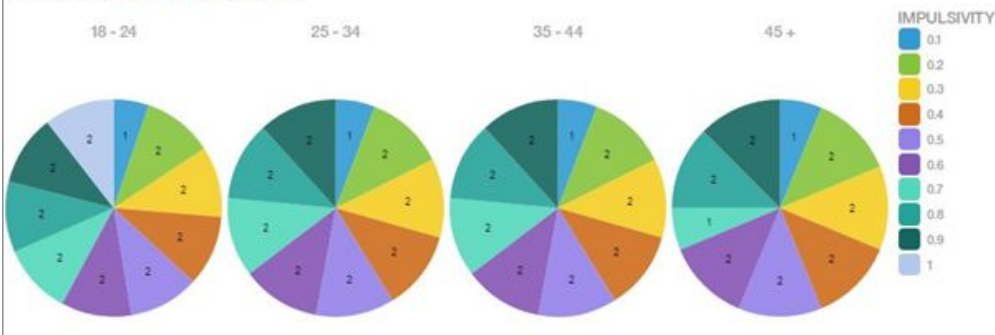
The results we found were impressive. For inactive users of coke, the age group 18-24 have the highest impulsivity among all the other age groups. It's almost the same for other age groups.

As this is a young age group, people from this age group tend to have high impulsivity, and the age group of 45+ has the lowest impulsivity as people from this age group tend to be older. For active users of coke, the impulsivity is almost same for all the ages, but the age group of 35-44 has the lowest impulsivity.

The difference of impulsivity between individuals who use coke and individuals who do not use coke is significant. People who use coke tend to have high impulsivity compared to the people who don't use coke across all age groups.

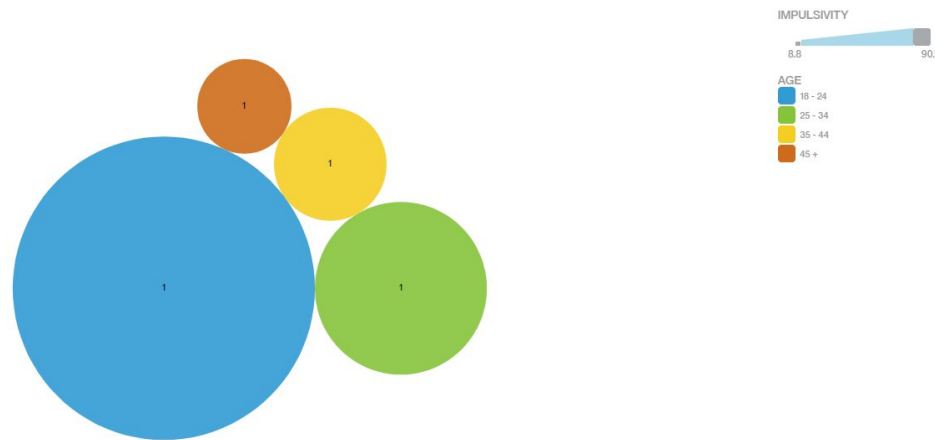
What is the breakdown of the number of **COKE** by **IMPULSIVITY** ?

Filtered by AGE: 4 selected



By observing the above pie chart, we can also suggest that people who have a high level of impulsivity are more exposed to coke across all age groups.

What are the values of **IMPULSIVITY** for each **AMPHET** and **AGE** ?
 Filtered by **AMPHET: 1**



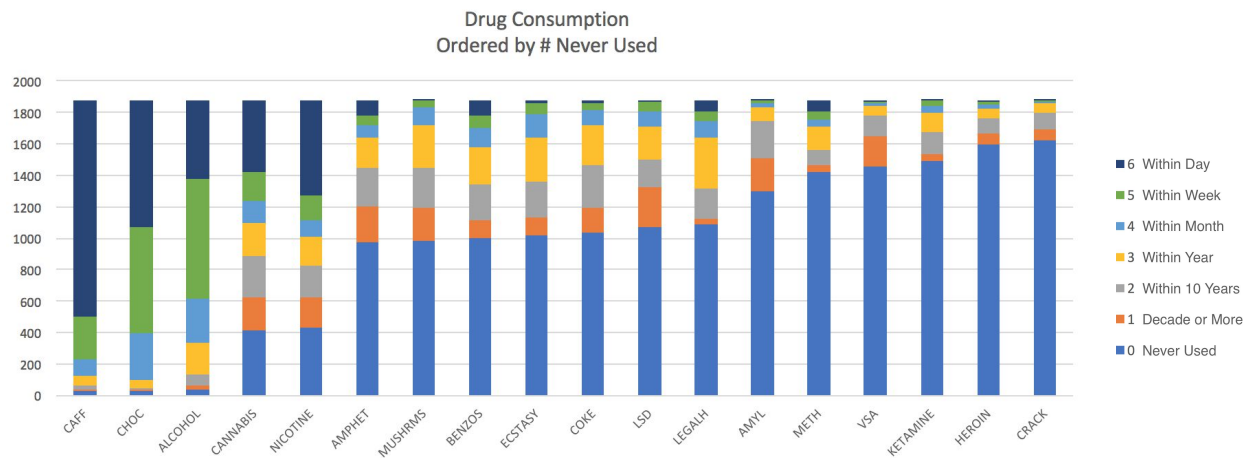
While non-cocaine users become less impulsive on average as they age, coke users do not follow this trend. This differs from most drugs, where impulsivity and abuse do not show such a strong correlation. For example, the bubble chart above shows that amphetamine users become less impulsive as they age, like most non-drug users. The underlying question now becomes: does prolonged cocaine use cause impulsivity, or does one's impulsivity cause them to continue to use the drug? We decided to leave this one up to the mental health professionals.

Data Analysis 3: What trends can we uncover in drug consumption rates by usage recency and by active users? What relationships are there between personality trait scores and drug users?

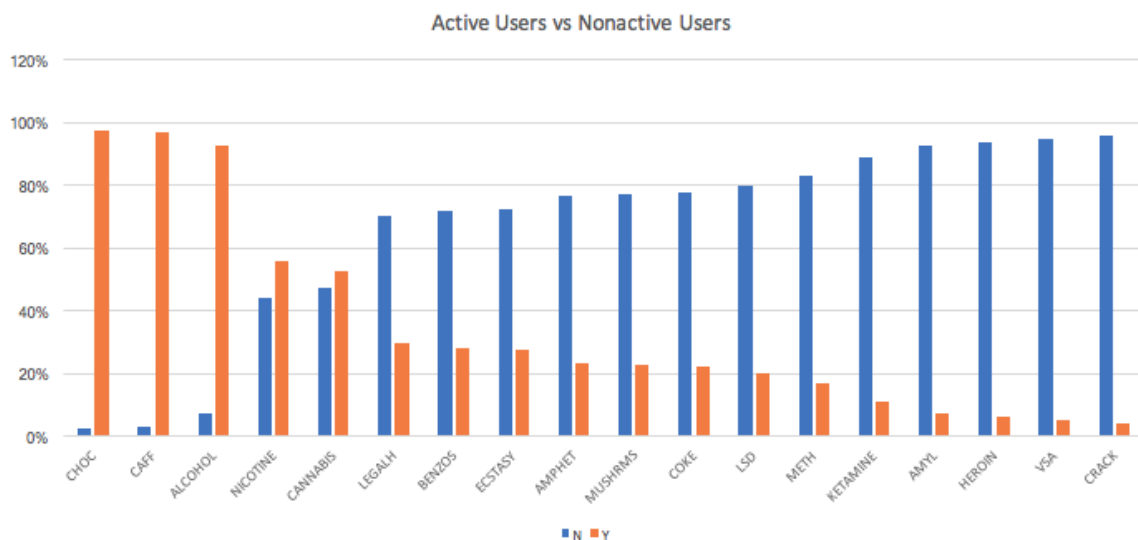
In order to find a social ranking of drug severity, we determined the usage rates for each drug, then ordered the drugs in terms of recency of use. In the below graph, drugs are ranked by number of people who have never tried the drug, with the assumption that few people will completely avoid very socially acceptable drugs and most people will completely avoid the least acceptable drugs. This scale ends up ranking caffeine as the most socially acceptable and crack as the least acceptable. Caffeine also has the highest rate of daily users, indicated by the long dark blue bar, and as you move to the right towards the more severe drugs, the daily usage rates generally fall, with some exceptions. The bumps in small daily usage rates tend to correlate with legal or prescription drugs taken daily by a small number of dedicated users. These include amphetamines (ie. Adderall), benzodiazepines (ie. Xanax), legal highs (ie. synthetic cannabinoids), and Methadone (opioid medication).

Possibly a sign of the times we are in, slightly more people had tried cannabis than nicotine, but nicotine consumption rates remained higher by every other measure. One interesting observation comes from alcohol consumption. While similar socially acceptable drugs have decreasing numbers from high daily users to low non-users, alcohol has a significantly higher weekly consumption rate than daily rate. This likely results from the social

acceptance of ‘weekend drinking’ versus ‘weekday drinking’. To illustrate, a quick Google search returns about 700 thousand results for “weekday drinking” as compared to over 85 million for “weekend drinking”.



To gather additional insights into usage trends, we grouped the usage data into Active versus Nonactive users. We defined Active users as those who had consumed the drug within the last year (usage levels 3-6) and Nonactive as those who had not consumed within the last year (usage levels 0-2). The graph below presents this new binary view of the consumption rates. This simplifies the data to show certain trends more clearly. Interestingly, the active user rate for alcohol is very close to that of chocolate and caffeine. All three fall in the 90th percentile, with alcohol only lagging about four percentage points behind the other two. Next we see that nicotine and cannabis rates are very similar, hovering around the 50 percent range with cannabis consumption about 3 percentage points below nicotine use. The next large group can be considered ‘Other Drugs’, made up of mostly recreational, illegal, and prescription drugs at or below 30 percent. We’ll look into some trends within this group in the next section.



Group	Drug	Group	Drug	Group	Drug
1	CAFF	3	AMPHET	4	AMYL
1	CHOC	3	MUSHRMS	4	METH
1	ALCOHOL	3	BENZOS	4	VSA
2	CANNABIS	3	ECSTASY	4	KETAMINE
2	NICOTINE	3	COKE	4	HEROIN
		3	LSD	4	CRACK
		3	LEGALH		

- Group 1: Caffeine, Chocolate and Alcohol are the highly consumed drugs by active users. Most people consume Caffeine daily, followed by eating Chocolate daily. Most people consume alcohol weekly. 90%+ consume Chocolate, Caffeine, and Alcohol actively.
- Group 2: Cannabis and Nicotine are the moderately consumed drugs by active and inactive users. About 50% of people consume Nicotine and/or Cannabis actively. Most of these people consume Nicotine daily, rather than weekly. Most of these people consume Cannabis daily, rather than weekly.
- Group 3: Amphetamine, Mushrooms, Benzodiazepines, Ecstasy, Cocaine, LSD and Legal Highs were consumed by Non-active users. Less than 30% of people consume Group 3 actively, and 52 to 58% of people never try Group 3.
- Group 4: Very few active users have consumed Amyl Nitrite, Methadone, VSA (Substance Abuse), Ketamine, Heroin, and Crack. Less than 17 % of people consume Group 4 actively, and about 70 to 86% of people never try group 4.

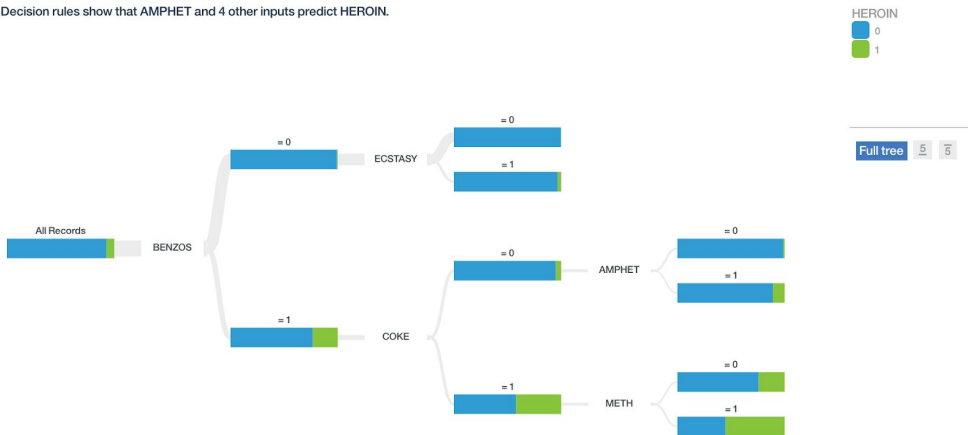
When we cross reference active users of each drug category with average scores of those users in seven personality traits, we begin to see relationships that describe the typical consumer in each category. First, average personality scores in the chocolate consumers group was exactly the same as our overall population averages, so we can use that group as a benchmark for comparison. As expected, our three most common drugs had very few differences in scores, with the notable exception of high neuroticism in caffeine consumers. Compared to consumers of other drugs, these consumers had relatively low neuroticism (N), openness to experience (O), impulsivity (I), and sensation seeking (SS) traits. They had relatively high agreeableness (A) and conscientiousness (C) scores. On the opposite end of the spectrum, active heroin users averaged the opposite trend. High neuroticism, impulsivity, and sensation seeking scores were most prominent in active users of volatile substances, crack, and heroin. Interestingly, the highest scores in openness to experience correlated with LSD and mushrooms users.

	CHOC	CAFF	ALCOHOL	NICOTINE	CANNABIS	BENZOS	COKE	AMYL	LSD	MUSHROOMS	LEGALH	ECSTASY	KETAMINE	AMPHET	METH	VSA	CRACK	HEROIN
Neuroticism	0.496	0.612	0.496	0.525	0.521	0.580	0.550	0.514	0.500	0.517	0.530	0.518	0.520	0.545	0.574	0.584	0.554	0.607
Openness to Experience	0.703	0.704	0.705	0.733	0.753	0.743	0.745	0.732	0.789	0.782	0.764	0.757	0.757	0.755	0.754	0.747	0.725	0.754
Extraversion	0.572	0.574	0.575	0.570	0.567	0.563	0.591	0.580	0.587	0.584	0.574	0.593	0.601	0.567	0.548	0.583	0.577	0.567
Agreeableness	0.644	0.643	0.644	0.628	0.626	0.611	0.602	0.602	0.631	0.622	0.624	0.623	0.630	0.620	0.601	0.601	0.601	0.563
Conscientiousness	0.614	0.612	0.614	0.579	0.575	0.565	0.567	0.581	0.593	0.573	0.563	0.576	0.572	0.550	0.558	0.541	0.535	0.541
Impulsivity	0.487	0.491	0.488	0.536	0.547	0.574	0.587	0.524	0.560	0.567	0.563	0.574	0.576	0.601	0.570	0.616	0.654	0.646
Sensation Seeking	0.557	0.563	0.564	0.644	0.677	0.688	0.715	0.655	0.718	0.713	0.709	0.717	0.712	0.717	0.696	0.734	0.762	0.741

Question 4: Can we use predictive models to determine drug consumption? If so, what are the main drivers of this prediction?

We ran a predictive model for a drug that shows the most extreme differences from our general population, but not the lowest usage rates. When running the a prediction tree for heroin use, we see that the largest predictors are in fact the use of other drugs. This gives is a strong predictive strength of 94%, with a 5 node tree. Use of benzodiazepines was the strongest predictor, with a predictive performance of .38. While our model predicted 80 occurrences of heroin use out of the sample of 995, it correctly predicted 44 individual cases. This gives us a specificity rate of 63%, but a much higher sensitivity rate of 96%. While this model would not be extremely reliable for identifying heroin users, it is very reliable in predicting nonusers. Noticeably absent is any reference to personality traits. What we can learn from this is that while there may be relationships between the use of particular drugs and trends in personality indicators, there is not a strong enough predictive link between the traits and drug use. Therefore, we cannot use personality traits and demographic data alone to predict an individual’s drug consumption risk accurately.

What is a predictive model for **HEROIN** ? (Predictive strength: 94%)
Decision rules show that AMPHET and 4 other inputs predict HEROIN.



Predictor importance			
Input	Value		
BENZOS	0.38		
COKE	0.30		
METH	0.25		
AMPHET	0.05		
ECSTASY	0.01		

	Predicted		Percent correct
	0	1	
0	889	36	96%
1	26	44	63%
	97%	55%	94%