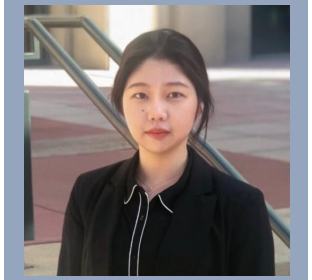
DataFest 2021

Identify and Predict Opioid Misuse with Demographic Profiles

Presented by **Team Odyssey**
Pratik Rath, Shiyan Chen, Wenshu Wu, Siddhartha Vanam
April 11, 2021

MEET THE TEAM

Pratik Rath
MEM '21
Duke University



Shiyan Chen
M.Eng FinTech
Duke University

Wenshu Wu
MQM
Duke University

Siddhartha Vanam
Economics & Statistics
UNC - Chapel Hill

01

THE NATIONAL OPIOID EPIDEMIC
Increasing Opioid misuse and overdose across the country

02

TOPIC INTERESTS
Demographic Information &. Opioids Use

03

HYPOTHESES & GOALS
Explanatory, Predictive Model

04

METHODS & MODELS
Data Handling, Tradeoffs of Models

05

CONCLUSIONS & FUTURE EXPLORATION
Reduce the survey completion time, better predictions of future opioid misuse.

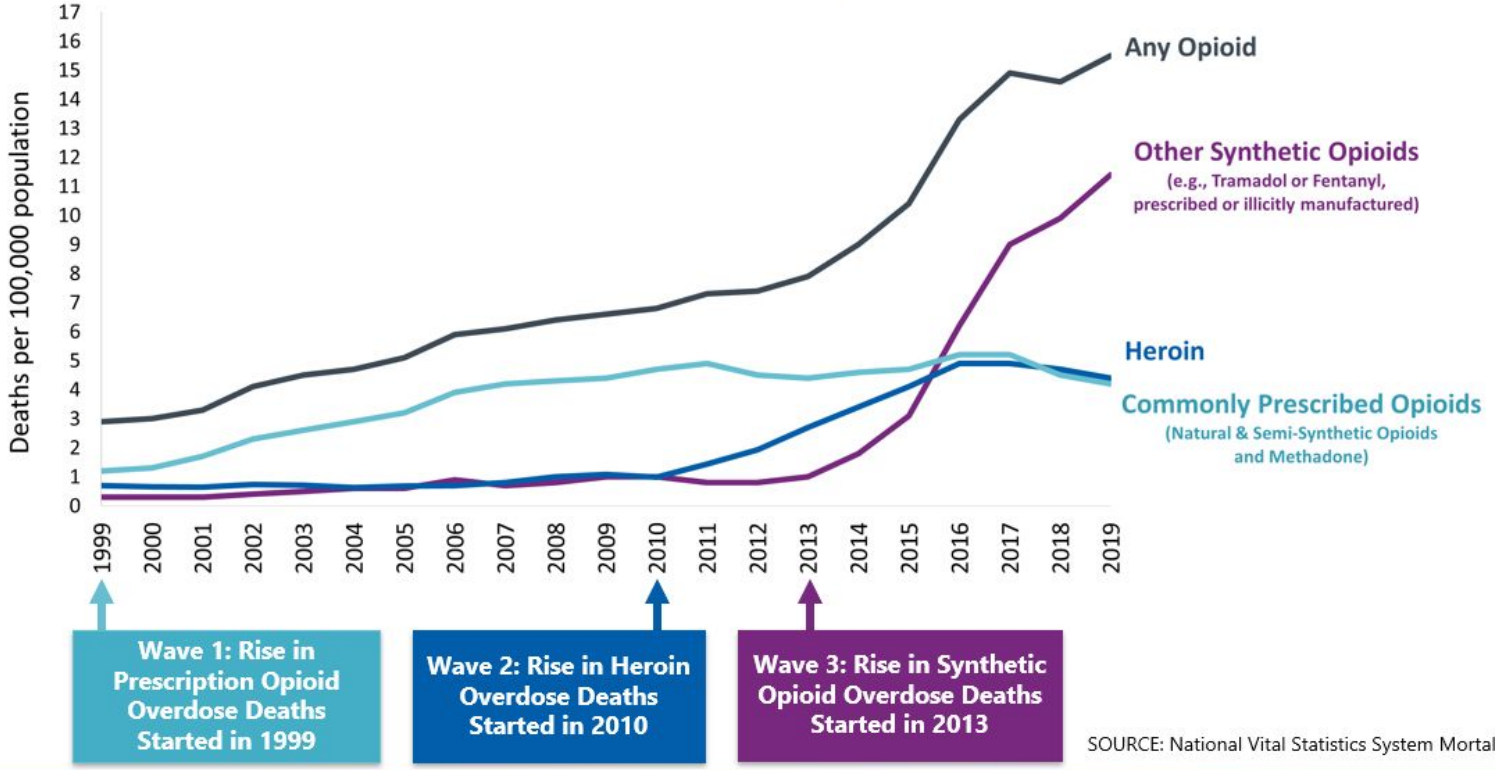
The National Opioid Epidemic

The annual cost of prescription opioid misuse in U.S alone is estimated to be **$78.5 billion.**

- **21-29%** patients misused prescription opioids.
- **8-12%** of the patients prescribed opioids for chronic pain management develop an opioid use disorder.
- **4-6%** of patients misusing prescriptions transition to heroin.
- **80%** of heroin users initially misused prescription opioids.

die every day from an opioid overdose
(including Rx and illicit opioids).

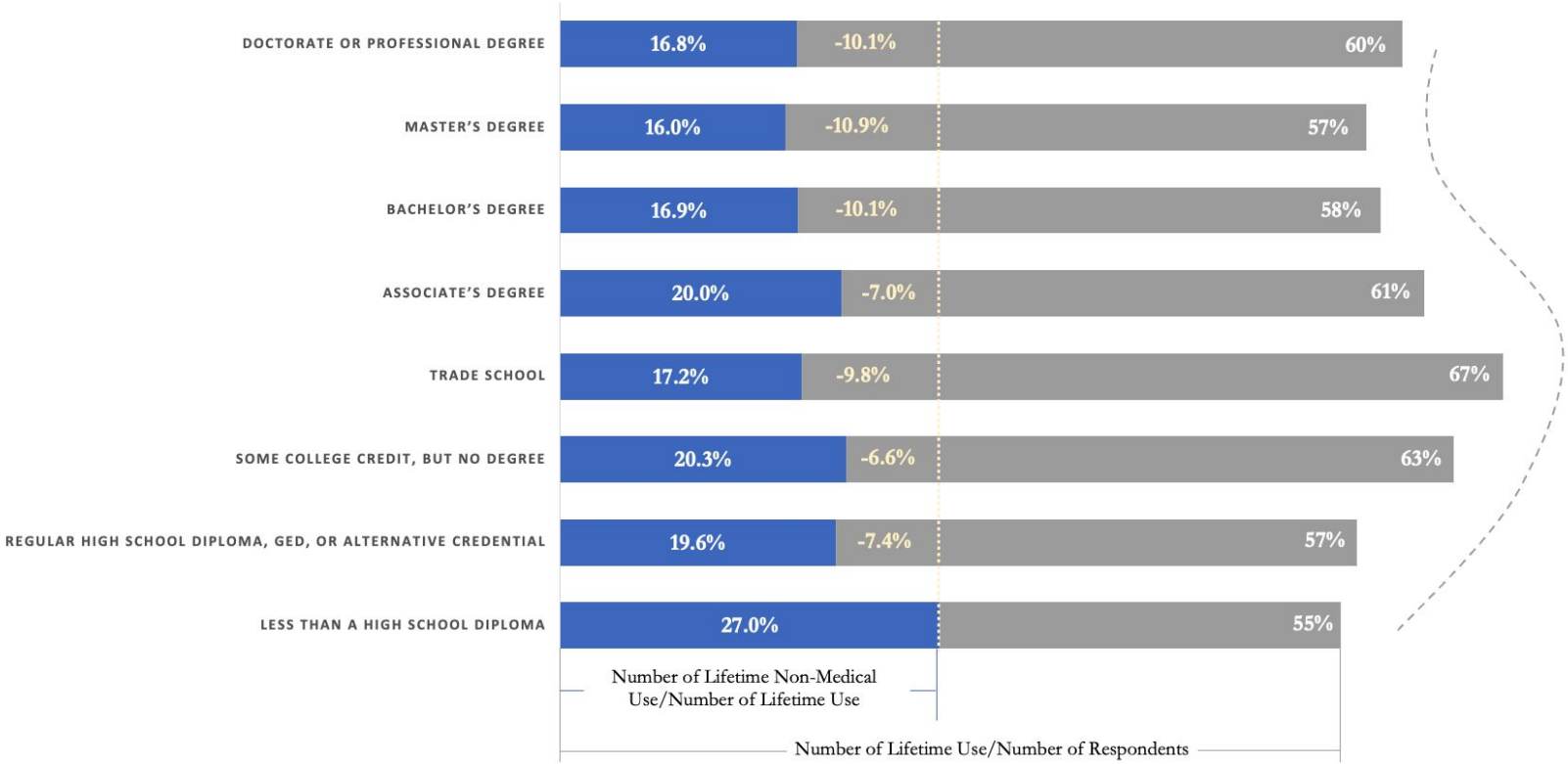
Three Waves of the Rise in Opioid Overdose Deaths



SOURCE: National Vital Statistics System Mortality File.

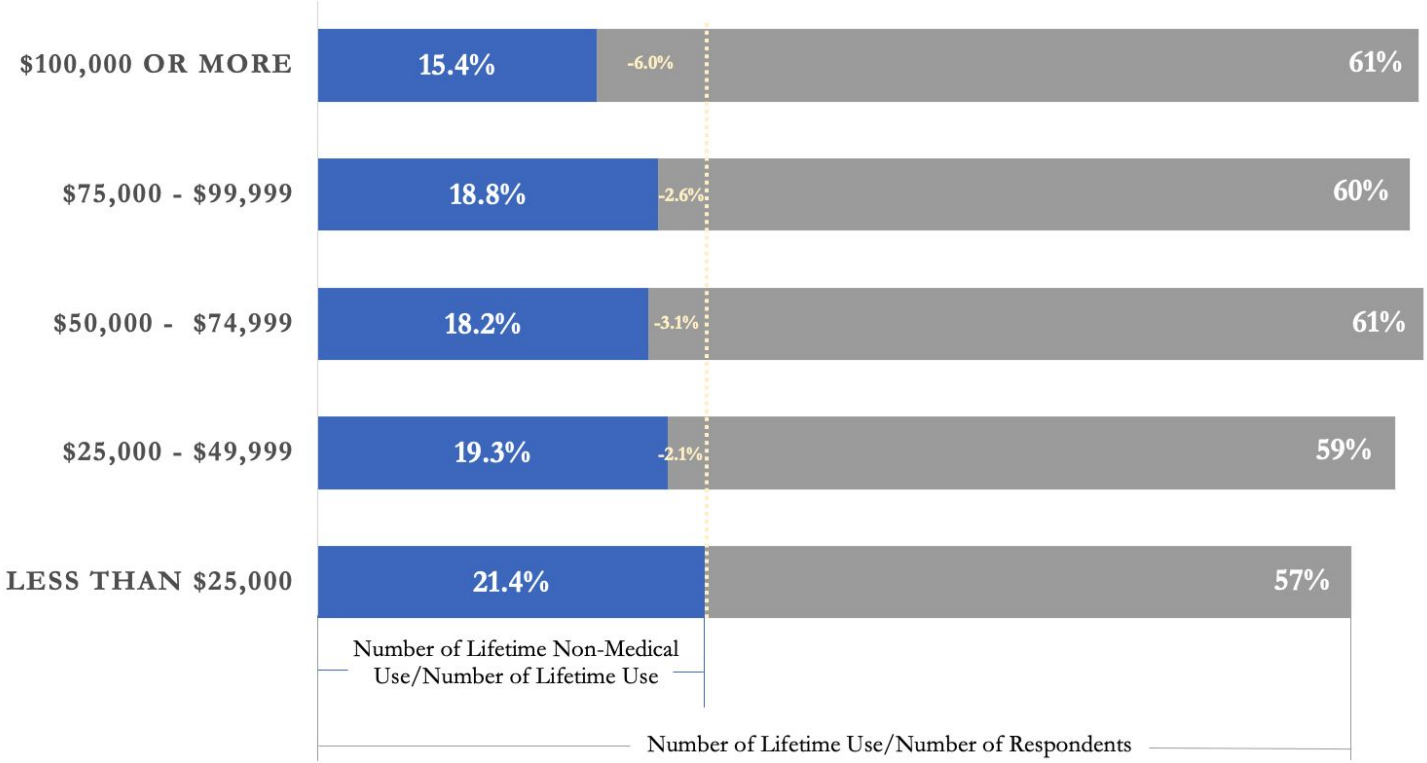
Education Level & Use of Opioids

A possible INVERTED U correlation between education level and lifetime use.
A possible NEGATIVE correlation between education level and the lifetime non-medical use.



Annual Household Income Level & Use of Opioids

A possible POSITIVE correlation between annual household income level and lifetime use.
A possible NEGATIVE correlation between annual household income level and the lifetime non-medical use.



GOALS & HYPOTHESES

GOALS

- **Explanatory**: Explore the effect of ***Education*** and ***Income*** on an individual's proclivity to abuse opioids?
- **Predictive**: Can we predict with high accuracy a survey respondents' opioid abuse tendencies?
- **Improve Survey Design**

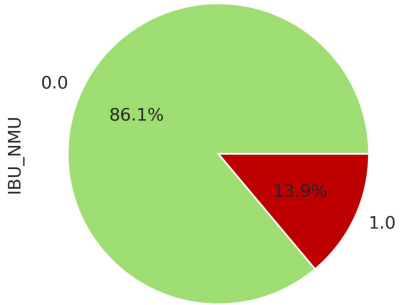
HYPOTHESES

- Education and Income will have a significant negative effect in opioid abuse
- Demographic Information will provide high predictive capability in identifying respondents that have abused opioids.

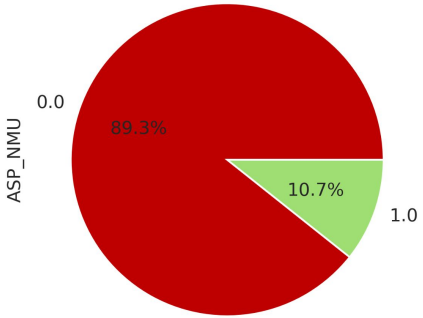
DATA DISCOVERY & ASSUMPTIONS

Distribution of Respondents that have Misused Non-Prescription Drug...

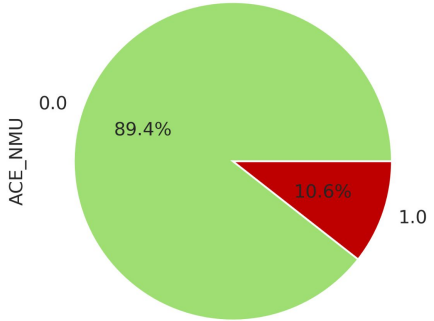
EVER



Ibuprofen

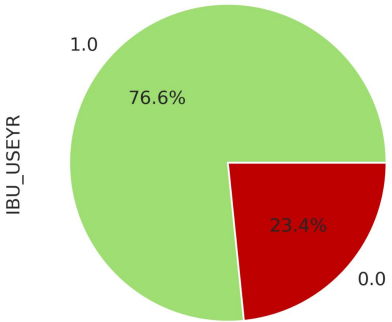


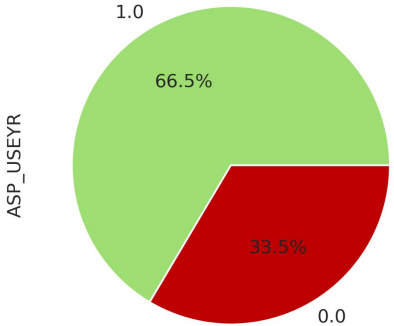
Aspirin

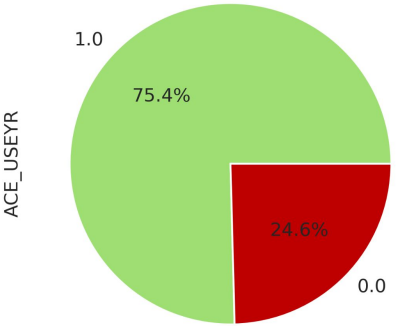


Acetaminophen

Last 12 Months







Model A vs. Model B

Logit Regression Results

Dep. Variable:	OP_NMU	No. Observations:	17923
Model:	Logit	Df Residuals:	17900
Method:	MLE	Df Model:	22
Date:	Sun, 11 Apr 2021	Pseudo R-squ.:	0.1750
Time:	02:25:12	Log-Likelihood:	-5101.8
converged:	True	LL-Null:	-6184.4
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.7244	0.155	-4.664	0.000	-1.029	-0.420
DEM_MALE	0.2247	0.054	4.197	0.000	0.120	0.330
DEM_AGE	-0.0291	0.002	-16.913	0.000	-0.032	-0.026
DEM_REGION_NE	-0.2674	0.083	-3.208	0.001	-0.431	-0.104
DEM_REGION_MW	-0.2172	0.081	-2.675	0.007	-0.376	-0.058
DEM_REGION_S	-0.0051	0.069	-0.073	0.942	-0.141	0.131
DEM_EDU	-0.0091	0.015	-0.591	0.554	-0.039	0.021
DEM_INCOME	0.0338	0.021	1.576	0.115	-0.008	0.076
DEM_RACE_AIAN	0.1556	0.178	0.873	0.383	-0.194	0.505
DEM_RACE_ASIAN	-0.5439	0.153	-3.553	0.000	-0.844	-0.244
DEM_RACE_BLACK	-0.0319	0.124	-0.257	0.797	-0.275	0.211
DEM_RACE_NHPI	-0.1861	0.365	-0.510	0.610	-0.901	0.529
DEM_RACE_WHITE	-0.2792	0.110	-2.539	0.011	-0.495	-0.064
DAST_1	1.1299	0.068	16.505	0.000	0.996	1.264
DAST_2	0.9381	0.131	7.170	0.000	0.682	1.195
DAST_3	-0.3389	0.062	-5.444	0.000	-0.461	-0.217
DAST_4	0.2509	0.122	2.049	0.040	0.011	0.491
DAST_5	0.5520	0.082	6.699	0.000	0.390	0.713
DAST_6	0.2930	0.132	2.226	0.026	0.035	0.551
DAST_7	0.1703	0.148	1.149	0.250	-0.120	0.461
DAST_8	0.4643	0.124	3.741	0.000	0.221	0.708
DAST_9	0.8459	0.106	8.017	0.000	0.639	1.053
DAST_10	-0.1516	0.156	-0.974	0.330	-0.457	0.153

Logit Regression Results

Dep. Variable:	OP_NMU	No. Observations:	17923
Model:	Logit	Df Residuals:	17867
Method:	MLE	Df Model:	55
Date:	Sun, 11 Apr 2021	Pseudo R-squ.:	0.3686
Time:	18:36:17	Log-Likelihood:	-3904.7
converged:	True	LL-Null:	-6184.4
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.9351	0.237	-3.953	0.000	-1.399	-0.471
ASP_NMU	0.5500	0.104	5.294	0.000	0.346	0.754
ACE_NMU	0.4069	0.102	3.972	0.000	0.206	0.608
IBU_NMU	0.5425	0.094	5.769	0.000	0.358	0.727
OXY_USE	0.4395	0.071	6.175	0.000	0.300	0.579
HYD_USE	0.6280	0.076	8.297	0.000	0.480	0.776
TRI_USE	0.6608	0.167	3.947	0.000	0.333	0.989
DRON_USE	0.5600	0.193	2.899	0.004	0.181	0.939
TRAM_USE	0.5009	0.073	6.849	0.000	0.358	0.644
ALP_USE	0.1143	0.080	1.423	0.155	-0.043	0.272
BAC_USE	0.1543	0.143	1.075	0.282	-0.127	0.436
OTH_RX_DRUG_USE	-1.3798	0.080	-17.188	0.000	-1.537	-1.222
DEM_HOME	0.0182	0.015	1.179	0.239	-0.012	0.049
DEM_MALE	0.3831	0.063	6.052	0.000	0.259	0.507
DEM_AGE	-0.0290	0.002	-12.364	0.000	-0.034	-0.024
DEM_REGION_NE	-0.1248	0.096	-1.298	0.194	-0.313	0.064
DEM_REGION_MW	-0.1537	0.093	-1.660	0.097	-0.335	0.028
DEM_REGION_S	-0.0652	0.080	-0.813	0.416	-0.222	0.092
DEM_EDU	-0.0039	0.018	-0.217	0.828	-0.039	0.032
DEM_INCOME	-0.0238	0.027	-0.884	0.377	-0.076	0.029
DEM_RACE_AIAN	-0.1671	0.208	-0.802	0.423	-0.576	0.241
DEM_RACE_ASIAN	-0.1975	0.175	-1.127	0.260	-0.541	0.146
DEM_RACE_BLACK	0.0705	0.143	0.495	0.621	-0.209	0.350
DEM_RACE_NHPI	-0.1812	0.398	-0.455	0.649	-0.961	0.599
DEM_RACE_WHITE	-0.4551	0.127	-3.580	0.000	-0.704	-0.206
DEM_INSUR	0.1194	0.069	1.724	0.085	-0.016	0.255
DEM_MARITAL	-0.0467	0.020	-2.354	0.019	-0.086	-0.008
DAST_1	0.5496	0.082	6.676	0.000	0.388	0.711
DAST_2	0.5756	0.157	3.673	0.000	0.268	0.883
DAST_3	-0.3135	0.075	-4.163	0.000	-0.461	-0.166
DAST_4	-0.0673	0.144	-0.468	0.640	-0.349	0.215
DAST_5	0.2592	0.096	2.697	0.007	0.071	0.448
DAST_6	0.1404	0.156	0.900	0.368	-0.165	0.446
DAST_7	0.2095	0.173	1.212	0.226	-0.129	0.548
DAST_8	0.2143	0.146	1.467	0.142	-0.072	0.501
DAST_9	0.3398	0.123	2.760	0.006	0.099	0.581
DAST_10	-0.5053	0.188	-2.684	0.007	-0.874	-0.136
COKE_USE	0.3223	0.101	3.194	0.001	0.125	0.520
KRAT_USE	0.3670	0.191	1.924	0.054	-0.007	0.741
SPEED_USE	0.1898	0.127	1.490	0.136	-0.060	0.440
METHAM_USE	0.2805	0.120	2.337	0.019	0.045	0.516
NPFENT_USE	0.5578	0.316	1.764	0.078	-0.062	1.177
LSD_USE	0.2304	0.106	2.183	0.029	0.024	0.437
BHYD_USE	0.2736	0.161	1.695	0.090	-0.043	0.590
BUP_USE	0.4385	0.141	3.106	0.002	0.162	0.715
COD_USE	1.0293	0.073	14.026	0.000	0.886	1.173
DIHY_USE	0.2164	0.176	1.230	0.219	-0.129	0.561
ELU_USE	0.2538	0.282	0.901	0.368	-0.298	0.806
FENT_USE	0.2453	0.111	2.211	0.027	0.028	0.463
GAB_USE	-0.3130	0.090	-3.484	0.000	-0.489	-0.137
KTM_USE	0.2698	0.157	1.716	0.086	-0.038	0.578
METH_USE	0.4955	0.126	3.944	0.000	0.249	0.742
MORPH_USE	0.1732	0.073	2.365	0.018	0.030	0.317
OXYM_USE	0.2266	0.136	1.668	0.095	-0.040	0.493
SUF_USE	0.7092	0.208	3.409	0.001	0.301	1.117
TAP_USE	0.6100	0.203	3.008	0.003	0.213	1.008

Model A vs. Model B

Logit Regression Results

Dep. Variable:	OP_NMU	No. Observations:	17923
Model:	Logit	Df Residuals:	17900
Method:	MLE	Df Model:	22
Date:	Sun, 11 Apr 2021	Pseudo R-squ.:	0.1750
Time:	02:25:12	Log-Likelihood:	-5101.8
converged:	True	LL-Null:	-6184.4
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.7244	0.155	-4.664	0.000	-1.029	-0.420
DEM_MALE	0.2247	0.054	4.197	0.000	0.120	0.330
DEM_AGE	-0.0291	0.002	-16.913	0.000	-0.032	-0.026
DEM_REGION_NE	-0.2674	0.083	-3.208	0.001	-0.431	-0.104
DEM_REGION_MW	-0.2172	0.081	-2.675	0.007	-0.376	-0.058
DEM_REGION_S	-0.0051	0.069	-0.073	0.942	-0.141	0.131
DEM_EDU	-0.0091	0.015	-0.591	0.554	-0.039	0.021
DEM_INCOME	0.0338	0.021	1.576	0.115	-0.008	0.076
DEM_RACE_AIAN	0.1556	0.178	0.873	0.383	-0.194	0.505
DEM_RACE_ASIAN	-0.5439	0.153	-3.553	0.000	-0.844	-0.244
DEM_RACE_BLACK	-0.0319	0.124	-0.257	0.797	-0.275	0.211
DEM_RACE_NHPI	-0.1861	0.365	-0.510	0.610	-0.901	0.529
DEM_RACE_WHITE	-0.2792	0.110	-2.539	0.011	-0.495	-0.064
DAST_1	1.1299	0.068	16.505	0.000	0.996	1.264
DAST_2	0.9381	0.131	7.170	0.000	0.682	1.195
DAST_3	-0.3389	0.062	-5.444	0.000	-0.461	-0.217
DAST_4	0.2509	0.122	2.049	0.040	0.011	0.491
DAST_5	0.5520	0.082	6.699	0.000	0.390	0.713
DAST_6	0.2930	0.132	2.226	0.026	0.035	0.551
DAST_7	0.1703	0.148	1.149	0.250	-0.120	0.461
DAST_8	0.4643	0.124	3.741	0.000	0.221	0.708
DAST_9	0.8459	0.106	8.017	0.000	0.639	1.053
DAST_10	-0.1516	0.156	-0.974	0.330	-0.457	0.153

89.4% Accuracy

22 Variables

90.9% Accuracy

55 Variables

Logit Regression Results

Dep. Variable:	OP_NMU	No. Observations:	17923
Model:	Logit	Df Residuals:	17867
Method:	MLE	Df Model:	55
Date:	Sun, 11 Apr 2021	Pseudo R-squ.:	0.3686
Time:	18:36:17	Log-Likelihood:	-3904.7
converged:	True	LL-Null:	-6184.4
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.9351	0.237	-3.953	0.000	-1.399	-0.471
ASP_NMU	0.5500	0.104	5.294	0.000	0.346	0.754
ACE_NMU	0.4069	0.102	3.972	0.000	0.206	0.608
IBU_NMU	0.5425	0.094	5.769	0.000	0.358	0.727
OXY_USE	0.4395	0.071	6.175	0.000	0.300	0.579
HYD_USE	0.6280	0.076	8.297	0.000	0.480	0.776
TRI_USE	0.6608	0.167	3.947	0.000	0.333	0.989
DRON_USE	0.5600	0.193	2.899	0.004	0.181	0.939
TRAM_USE	0.5009	0.073	6.849	0.000	0.358	0.644
ALP_USE	0.1143	0.080	1.423	0.155	-0.043	0.272
BAC_USE	0.1543	0.143	1.075	0.282	-0.127	0.436
TH_RX_DRUG_USE	-1.3798	0.080	-17.188	0.000	-1.537	-1.222
DEM_HOME	0.0182	0.015	1.179	0.239	-0.012	0.049
DEM_MALE	0.3831	0.063	6.052	0.000	0.259	0.507
DEM_AGE	-0.0290	0.002	-12.364	0.000	-0.034	-0.024
DEM_REGION_NE	-0.1248	0.096	-1.298	0.194	-0.313	0.064
DEM_REGION_MW	-0.1537	0.093	-1.660	0.097	-0.335	0.028
DEM_REGION_S	-0.0652	0.080	-0.813	0.416	-0.222	0.092
DEM_EDU	-0.0039	0.018	-0.217	0.828	-0.039	0.032
DEM_INCOME	-0.0238	0.027	-0.884	0.377	-0.076	0.029
DEM_RACE_AIAN	-0.1671	0.208	-0.802	0.423	-0.576	0.241
DEM_RACE_ASIAN	-0.1975	0.175	-1.127	0.260	-0.541	0.146
DEM_RACE_BLACK	0.0705	0.143	0.495	0.621	-0.209	0.350
DEM_RACE_NHPI	-0.1812	0.398	-0.455	0.649	-0.961	0.599
DEM_RACE_WHITE	-0.4551	0.127	-3.580	0.000	-0.704	-0.206
DEM_INSUR	0.1194	0.069	1.724	0.085	-0.016	0.255
DEM_MARITAL	-0.0467	0.020	-2.354	0.019	-0.086	-0.008
DAST_1	0.5496	0.082	6.676	0.000	0.388	0.711
DAST_2	0.5756	0.157	3.673	0.000	0.268	0.883
DAST_3	-0.3135	0.075	-4.163	0.000	-0.461	-0.166
DAST_4	-0.0673	0.144	-0.468	0.640	-0.349	0.215
DAST_5	0.2592	0.096	2.697	0.007	0.071	0.448
DAST_6	0.1404	0.156	0.900	0.368	-0.165	0.446
DAST_7	0.2095	0.173	1.212	0.226	-0.129	0.548
DAST_8	0.2143	0.146	1.467	0.142	-0.072	0.501
DAST_9	0.3398	0.123	2.760	0.006	0.099	0.581
DAST_10	-0.5053	0.188	-2.684	0.007	-0.874	-0.136
COKE_USE	0.3223	0.101	3.194	0.001	0.125	0.520
KRAT_USE	0.3670	0.191	1.924	0.054	-0.007	0.741
SPEED_USE	0.1898	0.127	1.490	0.136	-0.060	0.440
METHAM_USE	0.2805	0.120	2.337	0.019	0.045	0.516
NPFENT_USE	0.5578	0.316	1.764	0.078	-0.062	1.177
LSD_USE	0.2304	0.106	2.183	0.029	0.024	0.437
BHYD_USE	0.2736	0.161	1.695	0.090	-0.043	0.590
BUP_USE	0.4385	0.141	3.106	0.002	0.162	0.715
COD_USE	1.0293	0.073	14.026	0.000	0.886	1.173
DIHY_USE	0.2164	0.176	1.230	0.219	-0.129	0.561
ELU_USE	0.2538	0.282	0.901	0.368	-0.298	0.806
FENT_USE	0.2453	0.111	2.211	0.027	0.028	0.463
GAB_USE	-0.3130	0.090	-3.484	0.000	-0.489	-0.137
KTM_USE	0.2698	0.157	1.716	0.086	-0.038	0.578
METH_USE	0.4955	0.126	3.944	0.000	0.249	0.742
MORPH_USE	0.1732	0.073	2.365	0.018	0.030	0.317
OXYM_USE	0.2266	0.136	1.668	0.095	-0.040	0.493
SUF_USE	0.7092	0.208	3.409	0.001	0.301	1.117
TAP_USE	0.6100	0.203	3.008	0.003	0.213	1.008

CONCLUSIONS

Survey completion time can potentially be reduced from 15 minutes to 2 minutes on average.

- Save organizer's money in data entry and storage by limiting size of relevant questions
- Save respondents' time and increase their willingness to fill out the survey
- Enables preventative care predictions for opioid misuse

Potential Limitations of our final model:

- The inclusion of some statistically insignificant variables to prevent Omitted Variable Bias
- Did not account for weighted distributions
- May not be generalizable for the entire US Population, the scope can be limited in the audiences who took the survey and where they took it (i.e. clinic, hospital, etc.)

ASA DATA FEST 2021 @DUKE

Demographic Information & Opioid Misuse

89.4% Accuracy
Short Survey that can be done in 2 minutes

Thank You!

Team Odyssey

REFERENCES

"Opioid Crisis." *Official Web Site of the U.S. Health Resources & Services Administration*, 30 Nov. 2020, www.hrsa.gov/opioids.

"Other Drugs." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 26 Jan. 2021, www.cdc.gov/drugoverdose/data/otherdrugs.html.

"Understanding the Epidemic." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, 17 Mar. 2021, www.cdc.gov/drugoverdose/epidemic/index.html.

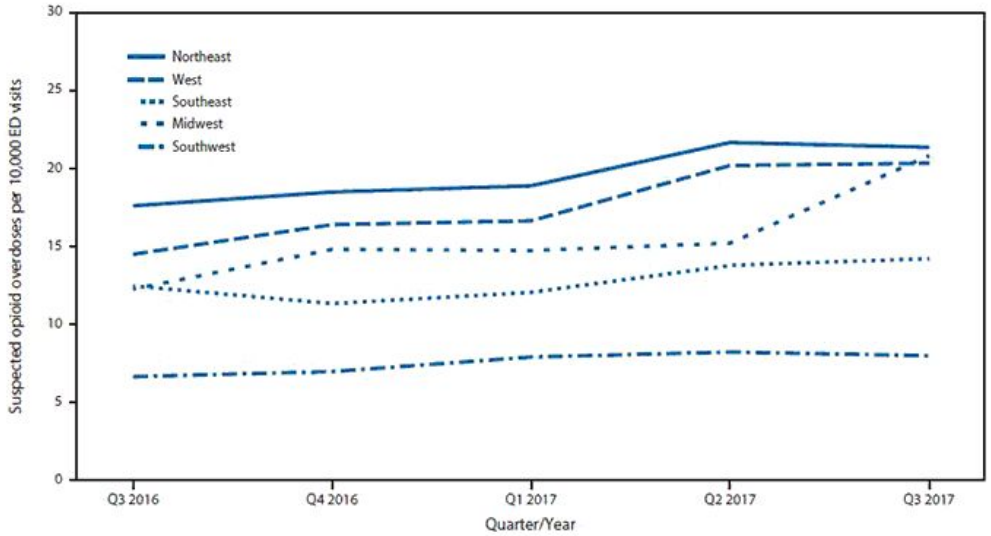
Gladden RM, O'Donnell J, Mattson CL, Seth P. Changes in Opioid-Involved Overdose Deaths by Opioid Type and Presence of Benzodiazepines, Cocaine, and Methamphetamine — 25 States, July–December 2017 to January–June 2018. MMWR Morb Mortal Wkly Rep 2019;68:737–744. DOI:http://dx.doi.org/10.15585/mmwr.mm6834a2

National Institute on Drug Abuse. "Opioid Overdose Crisis." *National Institute on Drug Abuse*, 11 Mar. 2021, www.drugabuse.gov/drug-topics/opioids/opioid-overdose-crisis.

Schuler, Megan S, et al. "Racial/Ethnic Differences in Prescription Opioid Misuse and Heroin Use among a National Sample, 1999–2018." *Drug and Alcohol Dependence,* 1 Apr. 2021, (https://www.sciencedirect.com/science/article/pii/S0376871621000831).

Appendix - Example of Weighted Feature

Opioid Overdoses by Region



Numbers of Respondents by Region