Analysis1Doc

# How much cannabis do Mexicans use

To the extent of the author´s knowledge, there is no rigorous effort to estimate the size of the cannabis market in Mexico. A study indicated that the surface of marijuana production eradicated from 2007 and 2015 was of 114,360 hectares, pointing Sinaloa as the major productor (36%), followed by Chihuahua (19.5%), and Durango (16.4%) (Resa Nestares, 2016). This number was taken then by the firm New Frontier data as the estimate for the total surface marijuana production in Mexico (New Frontier Data, n.d.). That same report estimated that Mexico had a potential market of 1.3 million of users and highlighted the economic potential of the Mexican production given the already existing price differential with the United States: “Currently, the average production cost of a pound of marijuana in legal producer countries is 400-800 dollars, 300-600 greenhouse and 200-300 dollars exterior…in Mexico de exterior production costs about 10 dollars”. Lastly, another company, Juicy Land, reported that the Mexican market might have a value of 250 million dollars and could create 45,000 employments (Garduño, 2021) although the methodology is not public.

The study will use a confirmatory approach to answer the questions: 1) What is the size of the cannabis market in Mexico in metric tons and profits? and 2) What is the average price paid by consumers? It will combine the use of survey data, following the methodology of studies developed at RAND such as Kilmer et al. did for the US (Kilmer et al., 2011) and Caulkins and Kilmer for the European Union. (European Commission. Directorate General for Justice., 2013).

# Running Code

First, we will upload the needed libraries

library("haven")  
library("survey")  
library("srvyr")  
library("gtsummary")  
library("data.table")  
library("forcats")  
library("plyr")  
library("dplyr")  
library("here")

## Preparing the data we need

We will set up the wd and the data. Let’s keep only the variables we will use and give them names that are easier to manage

# setwd("C:\Users\yperezd\Documents\GitHub\Quantitative-Analysis-Practicum-2024\student\_work\Sam PD")  
rm(list=ls(all=TRUE))  
  
#Load data  
encodat <- read\_dta(here("Data/ENCODAT\_2016\_2017\_Individual.dta"))  
encodat2 <- subset(encodat, select=c(id\_pers, entidad, desc\_ent, ds2, ds3, ds5a, ds9, ds16, di1a, di5a, di5a\_1, di6a, di8a, ponde\_ss))  
encodat2 <- as.data.table(encodat2)  
rm(encodat)  
  
#Recoding and relabeling variables  
setnames(encodat2, c("id", "state\_id", "state", "gender", "age" ,"indigenous", "educ", "occu", "ever", "times" ,"started\_py", "py" ,"pm", "weights"))  
  
#turning data into factors for the tables  
encodat2$ever <- factor(encodat2$ever)  
encodat2$age <- factor(encodat2$age)  
encodat2$times <- factor(encodat2$times)  
encodat2$started\_py <- factor(encodat2$started\_py)

We need the create age groups and regions to better analyze the information

#Preparing age groups  
encodat2$age2 <- as.numeric(levels(encodat2$age))[encodat2$age]  
breaks <- c(11, 21, 32, 43, 54, 66)  
encodat2$age2 <- cut(encodat2$age2, breaks=breaks)  
encodat2$age2 <- factor(encodat2$age2)  
  
#Preparing state regions  
central <- c(9, 12, 13, 15, 17, 20, 21, 29)  
west <- c(1,6,11,14,16,18,22,24,32)  
north <- c(2,3,5,8,10,19,25,26,28)  
southeast <- c(4,7,23,27,30,31)  
  
region\_lookup <- data.frame(  
 state\_code = c(central, west, north, southeast), # Combine state codes from all regions  
 region = c(rep("central", length(central)),  
 rep("west", length(west)),  
 rep("north", length(north)),   
 rep("southeast", length(southeast)))   
 )  
  
region\_lookup <- as.data.table(region\_lookup)  
region\_lookup[, state\_code := as.double(state\_code)]  
encodat2[, state\_id := as.double(state\_id)]  
encodat2 <- merge(encodat2, region\_lookup, by.x = "state\_id", by.y = "state\_code", all.x = TRUE)

Now, we need to rename the value labels

#Relabeling responses  
  
labels\_age2<- c("12-21", "22-32", "33-43", "44-54", "55-65")  
encodat2$age2<- as.double(encodat2$age2)  
encodat2[, age2 := mapvalues(age2, sort(unique(encodat2$age2)), labels\_age2)]  
  
labels\_ever<- c("Yes", "No", NA)  
encodat2$ever<- as.double(encodat2$ever)  
encodat2[, ever := mapvalues(ever, sort(unique(encodat2$ever)), labels\_ever)]  
  
labels\_started\_py<- c("Yes", "No")  
encodat2$started\_py<- as.double(encodat2$started\_py)  
encodat2[, started\_py := mapvalues(started\_py, sort(unique(encodat2$started\_py)), labels\_started\_py)]  
  
labels\_py<- c("Yes", "No")  
encodat2$py<- as.double(encodat2$py)  
encodat2[, py := mapvalues(py, sort(unique(encodat2$py)), labels\_py)]  
  
labels\_pm<- c("1-5 days", "6-19 days", "more than 20 days", "No")  
encodat2$pm<- as.double(encodat2$pm)  
encodat2[, pm := mapvalues(pm, sort(unique(encodat2$pm)), labels\_pm)]

Lastly, we set up the weights that are given in the data set

#Setting up the weights  
encodat\_w <- svydesign(id=~id, weights=~weights, data=encodat2)

## Analysis about marihuana use

Now, let’s look what the data is telling us

### We start with the question: Could you tell me if you have taken, used, tried Marijuana, also called “hashish”, “pot”, “coffee”, “yerba”, etc. to get high. We present statistics by the overall population, and dissagregated by gender, race and region

tbl\_svysummary(  
 include = c(ever),  
 label = list(ever ~ "Ever tried"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **N = 85,262,058** |
| --- | --- |
| Ever tried | 8.8% |
| Unknown | 311,011 |

tbl\_svysummary(  
 by = gender,  
 include = c(ever),  
 label = list(ever ~ "Ever tried"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **1**, N = 41,168,617 | **2**, N = 44,093,441 |
| --- | --- | --- |
| Ever tried | 14% | 3.8% |
| Unknown | 138,221 | 172,790 |

tbl\_svysummary(  
 by = age2,  
 include = c(ever),  
 label = list(ever ~ "Ever tried"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **12-21**, N = 23,372,489 | **22-32**, N = 21,538,196 | **33-43**, N = 16,045,431 | **44-54**, N = 13,888,546 | **55-65**, N = 10,417,396 |
| --- | --- | --- | --- | --- | --- |
| Ever tried | 9.7% | 12% | 8.2% | 6.4% | 4.9% |
| Unknown | 146,188 | 77,793 | 56,051 | 16,111 | 14,867 |

tbl\_svysummary(  
 by = region,  
 include = c(ever),  
 label = list(ever ~ "Ever tried"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **central**, N = 32,102,424 | **north**, N = 19,329,178 | **southeast**, N = 14,145,490 | **west**, N = 19,684,966 |
| --- | --- | --- | --- | --- |
| Ever tried | 8.0% | 9.7% | 8.1% | 9.8% |
| Unknown | 71,603 | 82,197 | 74,288 | 82,923 |

### How many times in your life have you used marijuana?

tbl\_svysummary(  
 include = c(times),  
 label = list(times ~ "# times"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **N = 85,262,058** |
| --- | --- |
| # times |  |
| 1 | 44% |
| 2 | 19% |
| 3 | 13% |
| 4 | 8.8% |
| 5 | 15% |
| Unknown | 77,887,863 |

tbl\_svysummary(  
 by = gender,  
 include = c(times),  
 label = list(times ~ "# times"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **1**, N = 41,168,617 | **2**, N = 44,093,441 |
| --- | --- | --- |
| # times |  |  |
| 1 | 42% | 52% |
| 2 | 18% | 21% |
| 3 | 14% | 11% |
| 4 | 9.2% | 7.4% |
| 5 | 16% | 8.8% |
| Unknown | 35,418,571 | 42,469,292 |

tbl\_svysummary(  
 by = age2,  
 include = c(times),  
 label = list(times ~ "# times"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **12-21**, N = 23,372,489 | **22-32**, N = 21,538,196 | **33-43**, N = 16,045,431 | **44-54**, N = 13,888,546 | **55-65**, N = 10,417,396 |
| --- | --- | --- | --- | --- | --- |
| # times |  |  |  |  |  |
| 1 | 38% | 42% | 51% | 52% | 53% |
| 2 | 23% | 19% | 16% | 14% | 12% |
| 3 | 14% | 15% | 9.4% | 12% | 15% |
| 4 | 8.6% | 8.7% | 10% | 9.9% | 5.5% |
| 5 | 17% | 15% | 13% | 12% | 15% |
| Unknown | 21,154,848 | 18,993,857 | 14,769,626 | 13,028,894 | 9,940,637 |

tbl\_svysummary(  
 by = region,  
 include = c(times),  
 label = list(times ~ "# times"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **central**, N = 32,102,424 | **north**, N = 19,329,178 | **southeast**, N = 14,145,490 | **west**, N = 19,684,966 |
| --- | --- | --- | --- | --- |
| # times |  |  |  |  |
| 1 | 46% | 41% | 48% | 43% |
| 2 | 20% | 17% | 17% | 19% |
| 3 | 16% | 11% | 14% | 12% |
| 4 | 8.4% | 11% | 6.5% | 8.6% |
| 5 | 9.1% | 20% | 14% | 17% |
| Unknown | 29,596,809 | 17,469,479 | 13,018,662 | 17,802,913 |

### Did you start using marijuana in the last 12 months?

tbl\_svysummary(  
 include = c(started\_py),  
 label = list(started\_py ~ "Started in the past year"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **N = 85,262,058** |
| --- | --- |
| Started in the past year | 14% |
| Unknown | 77,887,863 |

tbl\_svysummary(  
 by = gender,  
 include = c(started\_py),  
 label = list(started\_py ~ "started\_py"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **1**, N = 41,168,617 | **2**, N = 44,093,441 |
| --- | --- | --- |
| started\_py | 12% | 20% |
| Unknown | 35,418,571 | 42,469,292 |

tbl\_svysummary(  
 by = age2,  
 include = c(started\_py),  
 label = list(started\_py ~ "started\_py"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **12-21**, N = 23,372,489 | **22-32**, N = 21,538,196 | **33-43**, N = 16,045,431 | **44-54**, N = 13,888,546 | **55-65**, N = 10,417,396 |
| --- | --- | --- | --- | --- | --- |
| started\_py | 29% | 9.8% | 5.5% | 3.6% | 3.8% |
| Unknown | 21,154,848 | 18,993,857 | 14,769,626 | 13,028,894 | 9,940,637 |

tbl\_svysummary(  
 by = region,  
 include = c(started\_py),  
 label = list(started\_py ~ "started\_py"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **central**, N = 32,102,424 | **north**, N = 19,329,178 | **southeast**, N = 14,145,490 | **west**, N = 19,684,966 |
| --- | --- | --- | --- | --- |
| started\_py | 12% | 15% | 13% | 16% |
| Unknown | 29,596,809 | 17,469,479 | 13,018,662 | 17,802,913 |

### Have you used marijuana in the last 12 months?

tbl\_svysummary(  
 include = c(py),  
 label = list(py ~ "Used in the past year"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **N = 85,262,058** |
| --- | --- |
| Used in the past year | 13% |
| Unknown | 78,901,254 |

tbl\_svysummary(  
 by = gender,  
 include = c(py),  
 label = list(py ~ "Used in the past year"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **1**, N = 41,168,617 | **2**, N = 44,093,441 |
| --- | --- | --- |
| Used in the past year | 14% | 6.0% |
| Unknown | 36,109,374 | 42,791,880 |

tbl\_svysummary(  
 by = age2,  
 include = c(py),  
 label = list(py ~ "Used in the past year"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **12-21**, N = 23,372,489 | **22-32**, N = 21,538,196 | **33-43**, N = 16,045,431 | **44-54**, N = 13,888,546 | **55-65**, N = 10,417,396 |
| --- | --- | --- | --- | --- | --- |
| Used in the past year | 20% | 13% | 10% | 7.0% | 3.7% |
| Unknown | 21,799,738 | 19,243,097 | 14,839,749 | 13,059,706 | 9,958,964 |

tbl\_svysummary(  
 by = region,  
 include = c(py),  
 label = list(py ~ "Used in the past year"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **central**, N = 32,102,424 | **north**, N = 19,329,178 | **southeast**, N = 14,145,490 | **west**, N = 19,684,966 |
| --- | --- | --- | --- | --- |
| Used in the past year | 15% | 15% | 5.5% | 12% |
| Unknown | 29,897,761 | 17,740,961 | 13,163,401 | 18,099,130 |

### In the last 30 days, have you used marijuana?

tbl\_svysummary(  
 include = c(pm),  
 label = list(pm ~ "Used in the past month"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **N = 85,262,058** |
| --- | --- |
| Used in the past month |  |
| 1-5 days | 30% |
| 6-19 days | 9.5% |
| more than 20 days | 15% |
| No | 45% |
| Unknown | 83,440,437 |

tbl\_svysummary(  
 include = c(pm),  
 by = gender,  
 label = list(pm ~ "Used in the past month"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **1**, N = 41,168,617 | **2**, N = 44,093,441 |
| --- | --- | --- |
| Used in the past month |  |  |
| 1-5 days | 32% | 25% |
| 6-19 days | 11% | 3.6% |
| more than 20 days | 17% | 5.6% |
| No | 40% | 65% |
| Unknown | 39,747,560 | 43,692,877 |

tbl\_svysummary(  
 include = c(pm),  
 by = age2,  
 label = list(pm ~ "Used in the past month"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **12-21**, N = 23,372,489 | **22-32**, N = 21,538,196 | **33-43**, N = 16,045,431 | **44-54**, N = 13,888,546 | **55-65**, N = 10,417,396 |
| --- | --- | --- | --- | --- | --- |
| Used in the past month |  |  |  |  |  |
| 1-5 days | 29% | 32% | 28% | 31% | 32% |
| 6-19 days | 8.4% | 12% | 6.9% | 16% | 2.8% |
| more than 20 days | 13% | 13% | 27% | 19% | 26% |
| No | 50% | 43% | 38% | 34% | 40% |
| Unknown | 22,420,007 | 20,988,056 | 15,850,649 | 13,799,482 | 10,382,243 |

tbl\_svysummary(  
 include = c(pm),  
 by = region,  
 label = list(pm ~ "Used in the past month"),  
 statistic=list(all\_categorical()~"{p}%" ),   
 data=encodat\_w########  
)

| **Characteristic** | **central**, N = 32,102,424 | **north**, N = 19,329,178 | **southeast**, N = 14,145,490 | **west**, N = 19,684,966 |
| --- | --- | --- | --- | --- |
| Used in the past month |  |  |  |  |
| 1-5 days | 30% | 36% | 22% | 28% |
| 6-19 days | 9.2% | 12% | 5.8% | 8.3% |
| more than 20 days | 11% | 17% | 16% | 17% |
| No | 50% | 34% | 56% | 46% |
| Unknown | 31,477,251 | 18,823,049 | 13,946,253 | 19,193,884 |

This is the main variable that we are interested in, so let’s look at the distribution