**Srilasya Garigipaty**

**Data Analysis Report and Presentation**

**Student # 100822953**

**1.Executive Summary:**

The analysis is based on data from the Canadian Tobacco and Nicotine Survey (CTNS). The findings analyzed various age groups ranging from 15 years to 45 plus years to determine smoking habits, number of smokers across Canada, and determine the general patterns of smokers within the Canadian population. The data will inform Canadian policy when it comes to Tobacco use, and the findings from this analysis will provide a snapshot of tobacco use within various regions in Canada, as well as information related to smoking by population demographic. The analysis used data visualizations to determine key insights that could help create policy and campaigns against tobacco smoking and change unhealthy mechanisms that make it difficult to quit smoking.

**2.Introduction:**

The National Tobacco Strategy unit of Government of Canada wanted to gain information about the usage of tobacco across Canada, as there were many data gaps that policy makers did not know about. These data gaps include information such as the frequency of smokers within the country, the demographics of the smokers such as their age ranges and genders, the frequency they smoked, and information such as the number of times smokers have tried to quit, and the motives behind starting the nicotine usage. This report is created to help policy makers of the National Tobacco Strategy unit understand more about Tobacco use within Canada, and how to establish their policies and campaigns against smoking to have the highest beneficial outcomes.

**2.1 Background:**

The Canadian Tobacco and Nicotine Survey was developed in partnership with Heath Canada. The data used for the analysis portion of this report was collected in a survey format from 86000 respondents. The target population was non-institutionalized people aged 15 years and older living in Canada's provinces. The questionnaire included questions for respondents regarding their demographics, geographic regions respondents resided in, smoking habits such as when the respondent started smoking, how often they smoked, how many times they attempted to quit, the time frame it took them to quit, the reasons they started tobacco use, etc. This survey was initiated by The National Tobacco Strategy Unit Public Health Officer in collaboration with Health Canada Chief Officer to help collaborate and create policies and campaigns that would improve the Health, Lifestyle, and social conditions of Canadians by understanding how Tobacco is being consumed within the nation, and who it is being consumed by.

**2.2 Purpose of the Analysis:**

The analysis is being done to assist The National Tobacco Strategy Unit and Health Canada visualize their survey results. Since the data collected is large in number and has not utilized a linear methodology as some of the missing survey results were assumed, identifying the impacts and beneficiaries of the survey is not straightforward. The impacts of Tobacco use within Canada is not easily understood from survey results alone and it is beneficial to have a visual aid to assist in the understanding. Visualizing the smoking patterns in Canada allows for the identification of the impacts of smoking as well as who is being impacted the most. The findings from this analysis can be shared publicly through policy and campaigns to combat prevalent Tobacco use and improve the health of Canadians. The type of analysis being done is an ad-hoc analysis. This analysis is being done to address the specific business needs to target policies regarding tobacco use. Some questions being answered by this ad-hoc analysis include which gender smokes the most in Canada? What age group has the most prevalent number of smokers? What are the reasons different age groups start smoking? What age group attempts to quit the most, and what age groups have minimal attempts? Which province has the largest number of current smokers?

**2.3 Assumptions and Limitations:**

The assumption of the analysis include that the results produced by the survey and the data findings are generalized beyond the sample of 86000 respondents being studied. In addition, it is assumed that results of study will be meaningful to all stakeholders of this report. It is assumed that the participants selected for the survey are a good representation of the Canadian population.

Some limitations when interpreting the results of the analysis include that the some of the data is incomplete. Some result sections of the data lack information, and this has decreased the usability of some of the data. In addition, since the data is from surveys taken at a certain time in the respondent’s life, people do not always provide accurate information. The estimation of population characteristics from a sample survey is based on the premise that each person in the sample represents a certain number of other persons in addition to themselves. This number is referred to as the survey weight. The process of computing survey weights for each survey respondent involves several steps, and this statistical method may have errors in computation. Even though verification was done to ensure no duplicate records, and the data was cleaned up prior to starting the analysis, some out of scope and non-response records were remaining in the final data from the survey.

**2.4 Hypothesis:** There is no specific assumption or belief regarding the data.

**2.5 Questions**: The questions the data analysis must answer include the following:

* What is the percentage of smokers for each gender?
* What is the most prevalent smoking status amongst the nation?
* What pattern of smoking frequency is most common amongst various age groups?
* Which age group and gender smokes the least?
* Which age group and gender smokes the most?
* Which age group attempted to quit smoking the most?
* Which province has the largest percentage of total smokers?
* What are the various populations of the smokers and non-smokers by age group within Canada?
* What is the summary of key findings from the survey?

**2.6 Key Documents:**

**Data tables link**: https://www.canada.ca/en/health-canada/services/canadian-tobacco-nicotine-survey/2019-summary/2019-detailed-tables.html#t2

**Survey Description:** <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5305&dis=1>

**Full Study- Report by Statistics Canada:** <https://www.canada.ca/en/health-canada/services/canadian-tobacco-nicotine-survey/2019-summary.html>

**Key Definitions:** https://www.canada.ca/en/health-canada/services/canadian-tobacco-nicotine-survey/2019-summary/2019-definitions.html

**Tableau Analysis:** Attached along with this report.

**3. Methods**

The data source for this a analysis comes from the Canadian Tobacco and Nicotine Survey of 2019. This survey is conducted by The National Tobacco Strategy Unit in partnership with Health Canada and is a reliable data source because it is trustworthy, and a governmental entity. The analysis method includes first assessing the survey result tables. From these tables, it was determined whether there was enough data to answer some questions regarding smoking within Canada. Definitions for various terms in the survey were understood, and the data tables were cleaned up to be put in the right format for Tableau to make visualizations. The data quality is good. Quality assurance mechanisms were applied at all stages of the statistical process, and the survey data was validated and reviewed by the statisticians as a final verification of quality prior to release. Many validation measures were implemented, including:  
a. Verification of estimates through cross-tabulations  
b. Consultation with stakeholders internal to Statistics Canada  
c. Consultation with external stakeholders

Survey errors may have come from a variety of different sources. One error may be due to the sampling. Sampling error is defined as the error that arises because an estimate is based on a sample rather than the entire population. The 95% confidence interval values were considered to reduce this error in the final data.

There were a few missing values in the data. These values were omitted, and not considered for the analysis. The missing values were imputed with null in Tableau. During the analysis, these values were excluded from the charts when determining final insights, and only those categories where conclusions could be made because there were zero or few many missing values was considered.

**4.Results**

**4.1 Quality of the Data**

Data quality is referring to the assessment of the information in the survey, relative to its purpose. The data quality is measured according to five key metrics. These metrics include Accuracy, Consistency, Completeness, Integrity, and Timeliness. Since the data results are not changing in real time, the data is not 100% accurate. These survey results cannot be used to predict the smoking demographics in Canada for any other year other than 2019. The data is consistent. This consistency specifies that two data values pulled from separate data tables do not conflict with each other. The survey data was verified and has procedures in place before it was published. This indicates the data has Integrity. Lastly, the data is not timely. This means the data cannot be accessed based upon the business needs right away. Conducting and organizing a survey takes time, and the survey results of participant answers changes often. Data cannot be queried and retrieved for different time unless that survey has ample time to take place beforehand.

**4.2 Findings**

* **What is the percentage of smokers for each gender?**

The percentage of smokers for Men is 71% and Females is 11% at the 95% confidence interval. This indicates that smoking is more prevalent amongst men than women in Canada.

* **What is the most prevalent smoking status amongst the nation?**

The most prevalent smoking status in the nation is that the younger age groups from 15 to 19 have the highest percentage of individuals who have never smoked. As the age range increases, there is an increase in smokers. However, it can be noted that there are more individuals in the various age groups who have never smoked than those who have smoked.

* **What pattern of smoking frequency is most common amongst various age groups?**

Women were found to smoke less frequently than men, with men smoking more frequently monthly than women. As the age ranges increase, the percentage of women who do not smoke a lot is higher than men. Both genders seem to have majority of their gender stop smoking late in age, after 45 years.

* **Which** **age group and gender smokes the least?**

The age group that smokes the least is 15-19 years, and mostly women.

* **Which age group and gender smokes the most?**

The age group that smokes the most are around 25 to 45 years of age, and mostly men.

* **Which age group attempted to quit smoking the most?**

The age group that tried to quit smoking the most were mostly 45+, however there is a trend of individuals who are 25+ trying to quit smoking several times as well.

* **Which province has the largest percentage of total smokers?**

The province of New Brunswick has the largest percentage of total smokers, whereas of the province of British Colombia has the least.

* **What are the various populations of the smokers and non-smokers by age group within Canada?**

The following estimate populations were found of smokers and nonsmokers within various age groups in Canada.

15-19 Never Smoked: 35,035,552

20-24 Never Smoked: 30,571,707

25+ Never Smoked:22,244,199

25-44 Never Smoked:26,143,374

45+ Never Smoked:19,843,476

25+ Former Smokers: 10,578,186

25-44 Former Smokers: 6,414,431

45+ Former Smokers:13,166,465

25+ Current Smokers: 4,688,912

25-44 Current Smokers: 4,989,002

45+ Current Smokers: 4,501,355

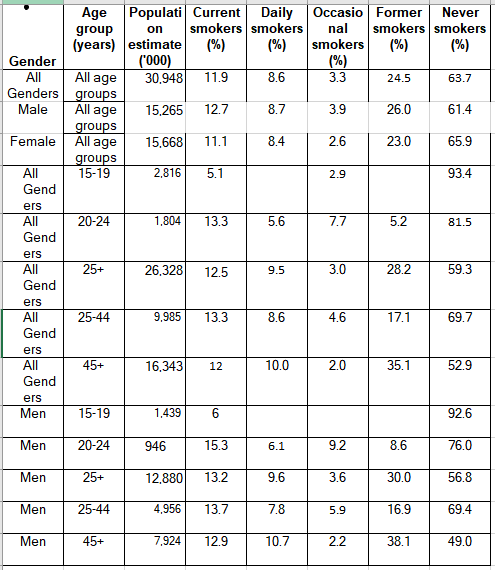
* **Summary of key findings?**

Based upon the survey results, it seems that younger and middle-aged individuals smoke the most compared to the elder age group. The gender of these individuals is prevalently male, however there is an increasing trend in smoking amongst females. New Brunswick has the largest percentage of smokers, followed closely by Manitoba and Nova Scotia. British Colombia has the least percentage of smokers. The elderly population 45+ attempts to quite smoking more than middle age, however there is an increasing pattern of attempts to quit by middle-aged individuals.

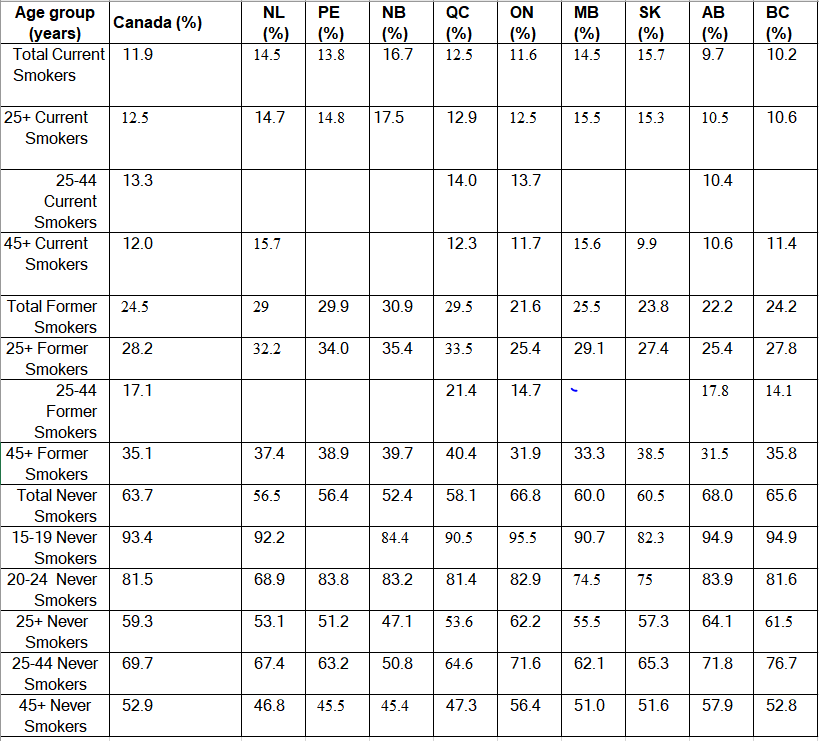
**Appendix**

**Cleaned up data tables by converting html page to Microsoft Excel and cleansing the data for analysis to create visualizations. Tables are from Data Tables link located Key Documents Section.**

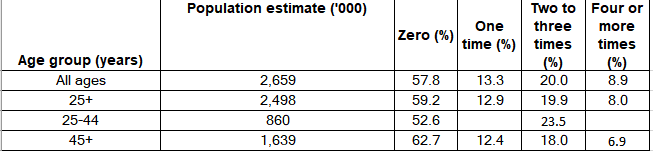
**Table 1. Smoking status by age group and gender, 2019**



**Table 2: Smoking status by province and age group, 2019**



**Table 3: Number of times attempted to quit smoking in the past 12 months among daily smokers, by age group, 2019**

 **Video Link:**

[**https://youtu.be/wKwSqMgkTcg**](https://youtu.be/wKwSqMgkTcg)

