Health App

Analyzing consumer's perception towards health applications

The objectives of our project was to assess the attitudinal and behavioral predictors of the use of health apps.

Executive Summary

Mobile phone use and the adoption of healthy lifestyle software apps ("health apps") are rapidly proliferating. There is limited information on the users of health apps in terms of their social demographic and health characteristics, intentions to change, and actual health behaviors. Over the past decades, health apps have seen significant progress and are mainly due to the growing usage of mobile phones. Presently, the number of mobile subscribers worldwide is expected to have extended to over 7 billion, of which 1 billion are smartphones with mobile broadband connections. Almost all developed and some developing countries have mobile penetration greater than 100%. The expectations are high form Health for both developed and developing regions in the world.

With the growing market demand for health apps, many businesses are trying to leverage the market potential. These businesses are ranging from private companies to governments, who wish to provide the community with an ability to self-care or get access to health advice easily. For businesses, it becomes important to understand the consumer's perception in utilizing these applications and this is where data analytics could help.

Problem Definition & Significance

The aim of this project is to necessarily answer the following main research question:

What are the stipulations that affect individuals' certitude in engagement or disengagement of health or wellness apps?

We are trying to cater to all the businesses which are willing/already in the health and wellness applications business. We will analyze the data from the Health Information National Trends Survey (2021), a nationally representative sample of adult people (18 years or older) in the U.S., and provide insights to the businesses on what could be their target market.

The reason this problem is interesting is that just looking at the market growth of the health applications it increases by 33% CAGR, and this could serve as a help to the community as well. Some of the statistics on the market dynamics for the health and wellness apps are as follows:

40% of physicians believe mobile health technologies can help reduce the number of visits to doctor's offices
93% of physicians believe that mobile health apps can improve patient's health
31% of surveyed organizations offer a specific app for the patients while **30%** are currently developing an app
Consumer health apps targeting wellness cover **2/3rd** of the health space
70% of mobile health practitioners rate diabetes highest for its market potential over the next 5 years

But with such an increase in the market potential of the health and wellness applications, the other reason that the consumers are hesitant in using the health applications is the security and privacy implications of these applications. Many people find it difficult to trust the internet to upload their private health-related details.

Prior Literature

1) Peng, W., Kanthawala, S., Yuan, S. et al. A qualitative study of user perceptions of mobile health apps. BMC Public Health 16, 1158 (2016). https://doi.org/10.1186/s12889-016-3808-0

   - ✓ Examines and qualitatively determines the design and content elements of health apps that facilitate or impede usage from the users' perceptive.
   - ✓ Among the participants, only 57% had health apps. Four types of reasons were identified by the participants explaining why they did not adopt or start to use health apps.
     1. Low Awareness of health apps
     2. Lack of app literacy
     3. Cost
     4. Lack of need for health apps

2) Jain, J., Udinia, P., & Sahoo, P. (2017). A qualitative study to analyze the pros-cons and consumer's perception towards mHealth apps. The Pharma Innovation Journal, 6, 43-48.

- ✓ This study has established that health management with using an app requires constant stimulation.
- ✓ There are many apps for health and wellness, but this found that most of the consumers are not aware of health app use. Among the participants these were the factors which were identified as follows:
    1. Age
    2. Income
    3. Occupation
    4. Education
    5. Awareness

3) Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A Privacy Assessment in Mobile Health Apps: Scoping Review JMIR Mhealth Uhealth 2020;8(7):e18868
    - ✓ This scoping review aims to understand how privacy is assessed for mHealth apps, focusing on the components, scales, criteria, and scoring methods used.
    - ✓ An important conclusion from this study is that there is a lack of analyses pertaining to the types of personal information collected by the apps.
    - ✓ Minimization is one of the principles of the GDPR, so a greater effort should be made to analyze whether apps gather more personal information than is necessary.
4) Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res. 2016 Oct 31;18(11):e287. doi: 10.2196/jmir.5692. PMID: 27806926; PMCID: PMC5295827.
    - ✓ Examines the effectiveness of mobile phone apps in achieving health-related behavior change in a broader range of interventions
    - ✓ The studies suggest that some features improve the effectiveness of apps, such as less time consumption, user-friendly design, real-time feedback, individualized elements, detailed information, and health professional involvement.
5) Carroll, J. K., Moorhead, A., Bond, R., LeBlanc, W. G., Petrella, R. J., & Fiscella, K. (2017). Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. Journal of medical Internet research, 19(4), e125. https://doi.org/10.2196/jmir.5604

- ✓ Describes the sociodemographic characteristics associated with health app use in the recent US nationally representative sample and assesses the attitudinal and behavioral predictors of the use of health apps for health promotion.
- ✓ Used multivariable logistic regression models to assess sociodemographic predictors of mobile device and health app use and examine the associations between app use
- ✓ Results show that main users of health apps were younger individuals, had more education, reported excellent health, and had a higher income

6) Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, Carey JL, Hayes RB, Ding EY, Bennett GG, Pagoto SL Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps JMIR Mhealth Uhealth 2017;5(12):e190

   - ✓ This paper summarizes methods for evaluating the content, usability, and efficacy of commercially available health apps.
   - ✓ Researching commercial apps may include evaluation of the technical functions of the app, developer transparency, and policies regarding user data privacy and security.
   - ✓ Results suggested that requested permissions often surpassed what the app needed, which means these apps could pose an unnecessary threat to user privacy and safety.

Data Source/Preparation

The data which we used is from Health Information National Trends Survey 2021(HINTS). It had 350 survey questions and 3866 records. Our dataset majorly had healthcare-related questions based on a person's health and his knowledge about any healthcare application. A question's response was measured in leveled categories.
Our dependent variable was whether a person has used health wellness apps in the past 12 months. We used this as our dependent variable as we wanted to check the impact of the pandemic and other factors, as we know most of the doctor visits could be treated by some generic medicines or just by a video call appointment. Hence, we wanted to study this more and gain more insights into it.

Our independent variables were as follows:

Have.device	Whether a person has a device with which he can use to access the health and wellness applications
Pandemic	Whether the survey was taken during a pandemic or not.
Willing.share.Data	Whether the person is willing to share his personal medical information through any health app.
Health.Insurance	Whether the patient has health insurance or not.
NotAccessed_ConcernedPrivacy	Whether the patient is concerned about his privacy and sharing information online.
Smoke	What time of a smoker the person is?
Age	Age of the person taking the survey
BMI_CDC	To check whether the person is fit or not.
Lifestyle	To check what kind of lifestyle the person has.

Gender	Gender of the person taking the survey.
HasaPartner	Whether the person has a partner or not.
Incomeranges	Whether the person belongs to lower, middle, or upper class
AlcoholIntake	Based on the number of drinks how much alcohol does the person consume in a day?
Health_cat	What health category does a person belong to.
Race	Which race does the person belong to?

Cleaning and Processing:

We had almost 350 survey questions, we have combined most of the similar questions and checked their correlation with each other, we successfully crunched it down to 105 questions, we used IBM SPSS software for this process. After this, we have converted many continuous variables to labeled categorical such as BMI_CDC was calculated from Height & Weight. We have many other variables which were calculated similarly.

List of calculated variables

Variable Name	Formula	explanation
BMI_CDC	BMI=Wt in Kg/height in m^2	BMI categories would give a clearer picture compared to just weight and height
Lifestyle	Based on no of days a person is physically active	Categories would be more appropriate while comparing.
Alcohol_Intake	Based on no of drinks per day	
Health_cat	Based on the fitness score rated in the survey	

Medical_conditions_cat	Based on the information about persons prior disease	
Race	Combined multiple races into 5 general categories – White, African American, Asian, Hispanic, Pacific Islanders.	
Smoker	Classified as a type of smoker based on no of cigarettes per day	
IncomeRanges	Classified into Upper, Lower, and Middle classes.	

Variable choice

What are your predictors of interest? Explain the rationale for these predictors.

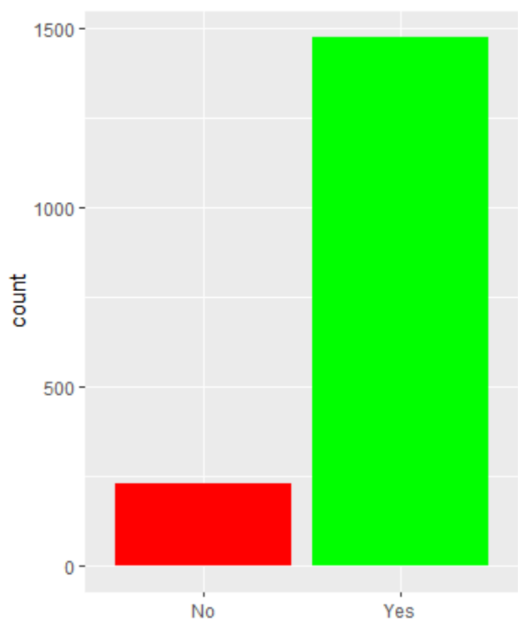
Variable Name	Effect	Rationale
BMI_CDC	+	Overweight/Obese tend to use heath/fitness apps more than healthy people.
Lifestyle	+	A person with an active lifestyle should be healthier than one with sedentary and should use these applications more
Alcohol_Intake	+	A person with High-level alcohol intake should use health applications more than compared to a non-drinker
Health_cat	-	People with poor health should use it more compared to healthier people

Medical_conditions_cat	-	People with heart conditions/lung cancer will be using it more than a person without any disease
Race	No Effect	There should not be any racial bias
Smoker	+	Chain smokers should use it more than non-smokers
IncomeRanges	-	People belonging to the lower and middle class should use it more as it would be a cheaper option.
Caregiving	+	Caregiving should use it more as it would be easier for him to access it remotely
HasPartner	No Effect	No effect
Age	+	Older people should use it more

Descriptive Analysis & Data Visualizations

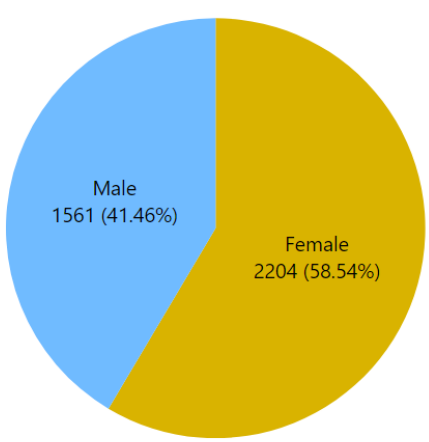
What patterns/trends do you see in your data? What do you infer from these trends?

Dependent variable: About 55% of users say they use Health apps in our dataset.

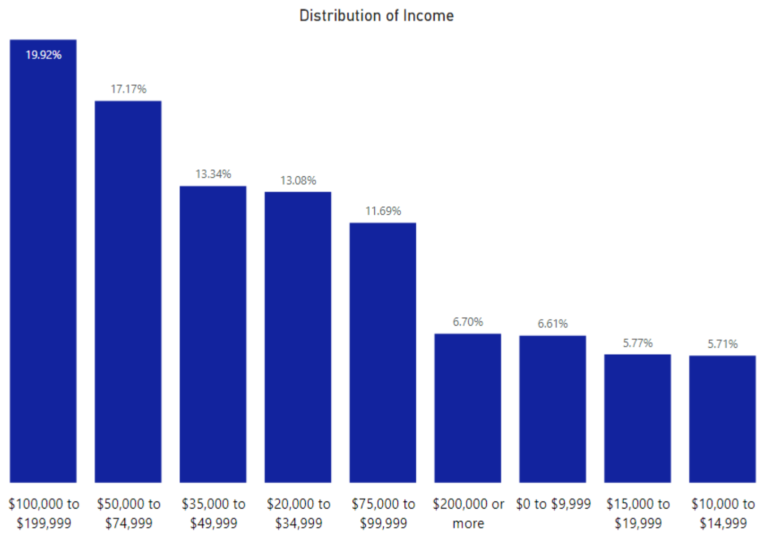


Other Visualizations:

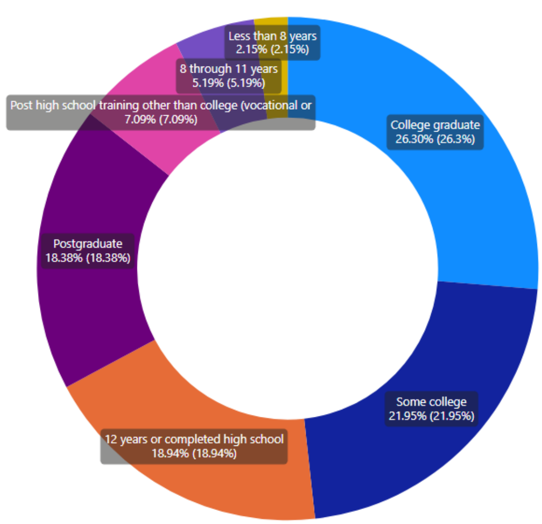
Gender: The data has about 58% females who participated in the survey as compared to 42% males.



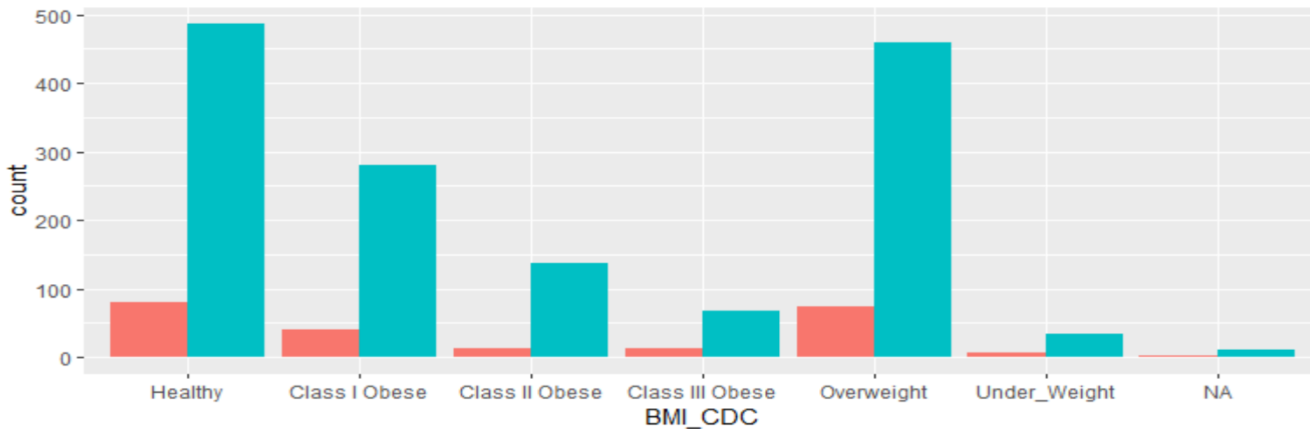
Income Ranges: The data has answers from most people with incomes higher than 40K. Suggesting high-income people use health apps more.



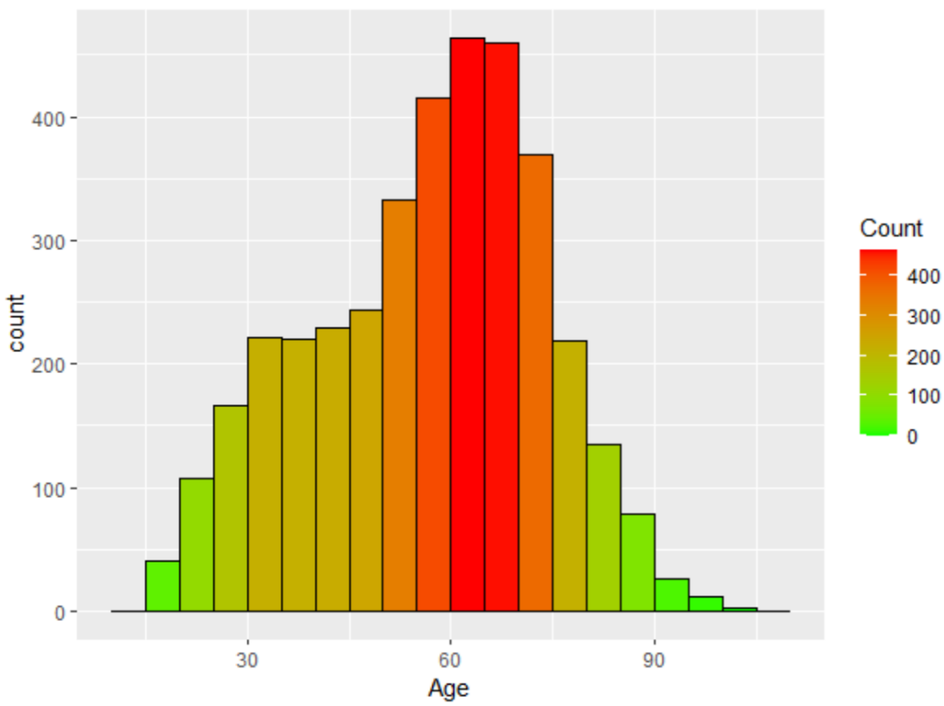
Education: The data has answers from around 26.3% of college graduates followed by 22% from some colleges. The lowest share comes from less than 8 years of education.



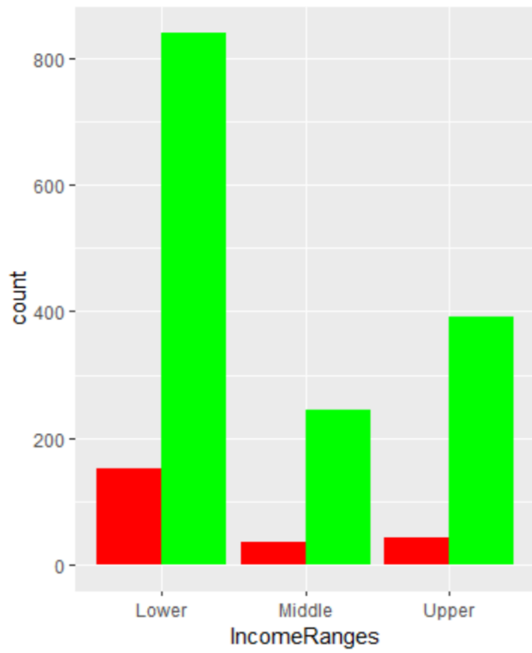
BMI_CDC: healthy people tend to use these applications more than other categories.



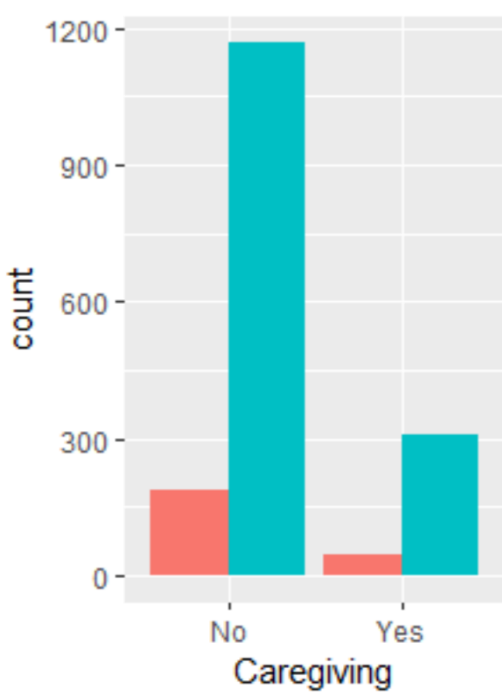
Age: distribution chart shows that more people are within the 40-80 age range and the majority people are from 60-70 Age.



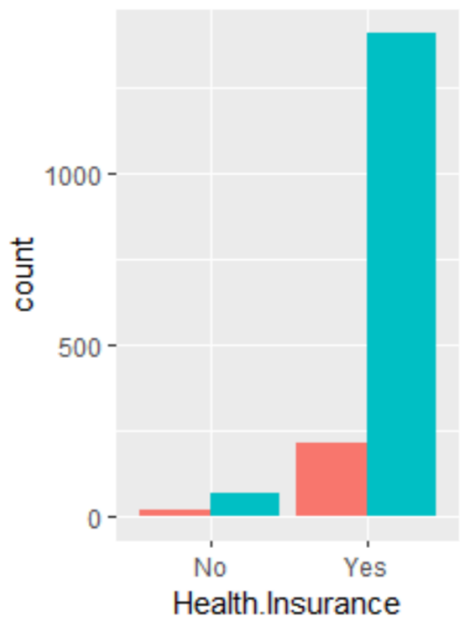
Income Ranges: people in lower-income ranges tend to use it more.



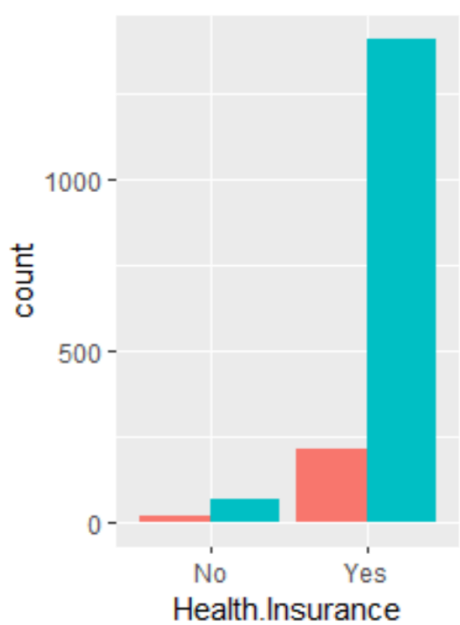
Caregiving: people who are not caregiving use it more than caregiving.



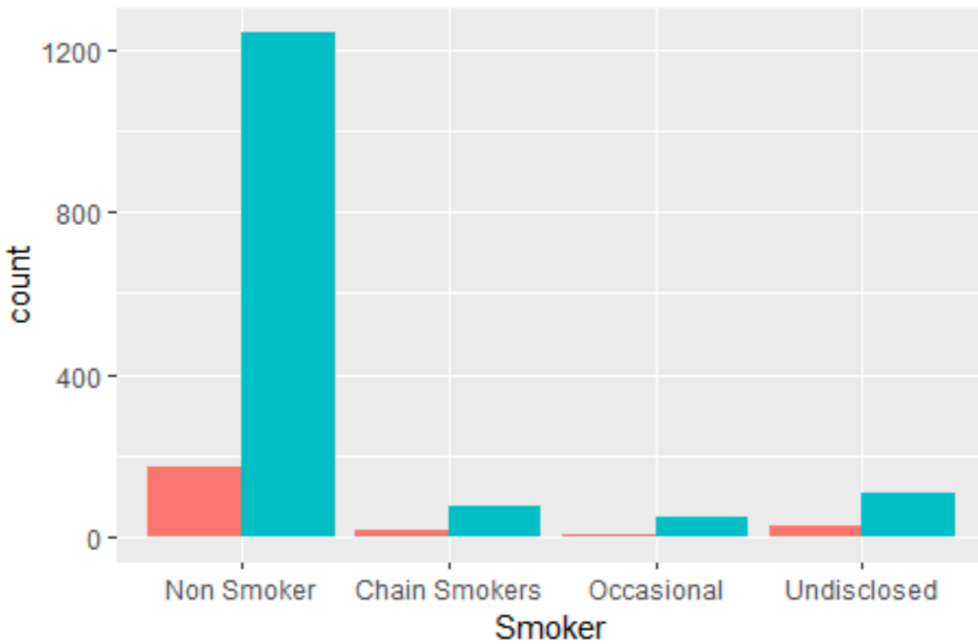
Health Insurance: people who have health Insurance use it more than people without health insurance.



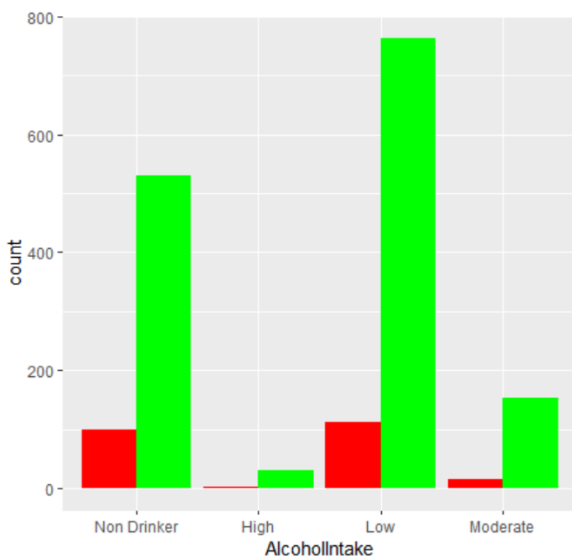
Gender: Females use it more than males.



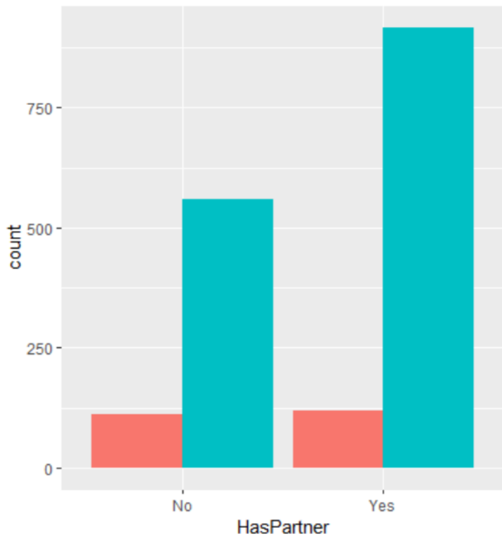
Smoker: Non-Smokers use it more than smokers.



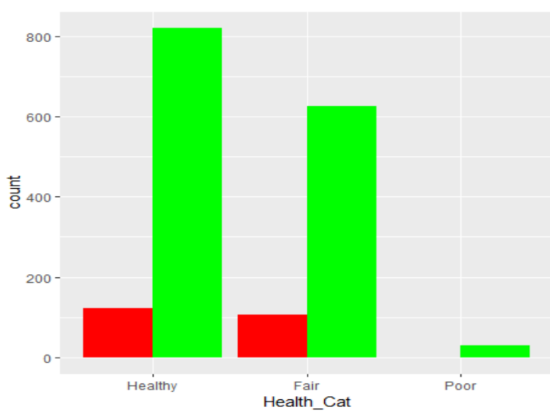
Alcohol Intake: Nondrinkers and low intake people use it more.



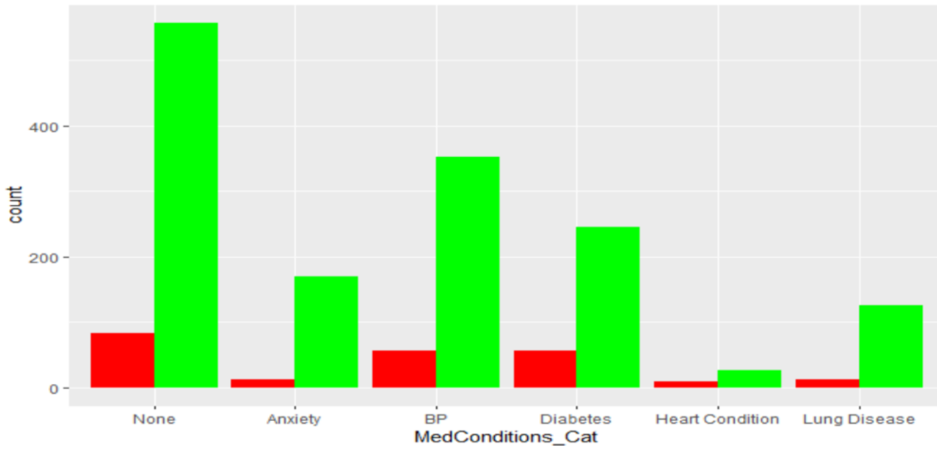
Has Partner: people with a partner use it more.



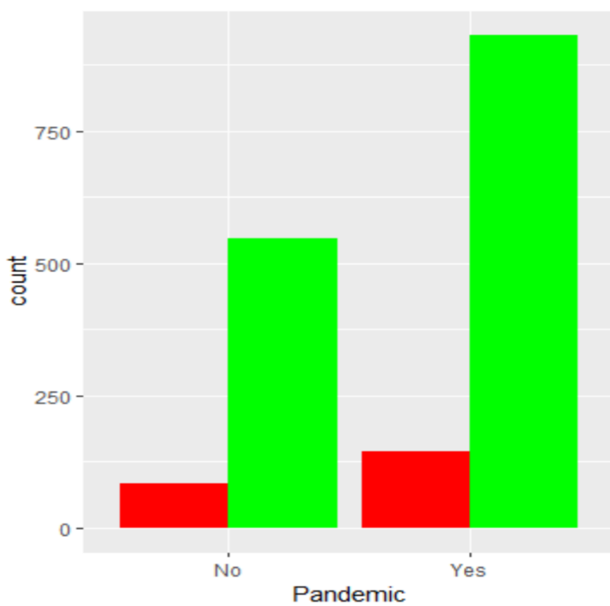
Health Category: Healthy people use it more than other categories.



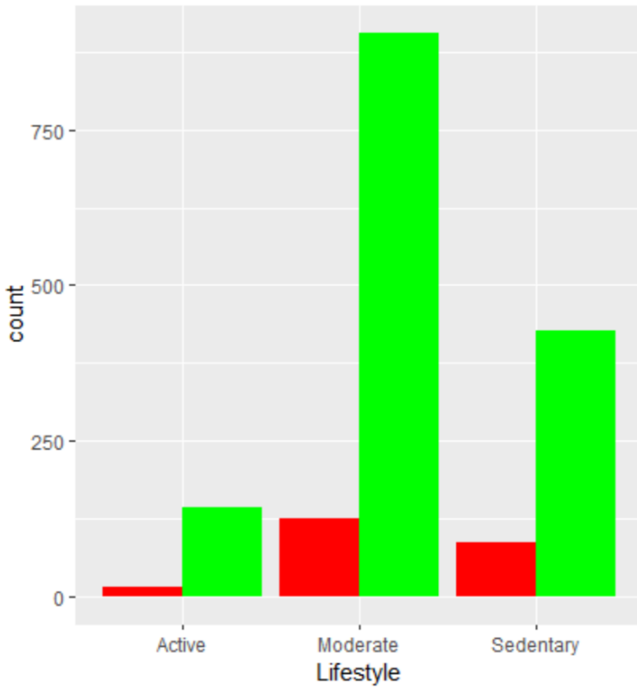
Medical Conditions: people without medical conditions use it more than people with a condition.



Pandemic: People have started using it more during the pandemic.



Lifestyle: people having a moderate lifestyle use it more than a sedentary lifestyle.



Models

The dependent variable we have is Used.Health.Wellness.Apps (whether you use a health wellness app or not?) And the dependent variable here is binary i.e., it can have two outcomes, 0 (as in the person does not use the health wellness app) or 1 (which is the person uses the health wellness app). And when you have a dependent variable that is binary selecting the correct model becomes very important. Because we cannot use OLS models as the OLS model with the binary variable will fail most of the OLS assumptions and also with the OLS model we have a problem that is the fitted values do not fit the Y variable.
So to fit the Y variable with the fitted values we have to convert the binary Y variable into a continuous probability distribution function which will vary between 0 and 1 and by doing that what we are doing is that instead of fitting the Y variable on a straight line we are trying to fit the Y variable on an S-shaped curve and this is possible by using models like Logit or Probit and those are the models we are using here.

Logit :

```
m1=glm (Used.Health.Wellness.Apps ~ have.device + Health.Insurance
    +Caregiving+BMI_CDC+ Smoker+Age+Gender
    + HasPartner + Race + IncomeRanges +MedConditions_Cat+Health_Cat+Lifestyle
    +Pandemic + AlcoholIntake,data = dfb, family = binomial(link = "logit"))
```

Probit :

```
m2= glm (Used.Health.Wellness.Apps ~ have.device + Health.Insurance
    + Caregiving + BMI_CDC + Smoker+ Age+ Gender
    + HasPartner + Race + IncomeRanges+MedConditions_Cat+Health_Cat+Lifestyle
    +Pandemic+ AlcoholIntake, data = dfb, family = binomial(link = "probit"))
```

Comparing the Models – Stargazer:

	Dependent variable:	
	Used.Health.Wellness.Apps	
	logistic	probit
	(1)	(2)
have.deviceNone	-2.763*** (0.694)	-1.613*** (0.405)
have.deviceSmartphone	-0.734*** (0.162)	-0.394*** (0.089)
have.deviceTablet/Computer only	-0.794** (0.385)	-0.435* (0.223)
Health.InsuranceYes	0.352 (0.322)	0.167 (0.179)
CaregivingYes	0.196 (0.201)	0.100 (0.107)
BMI_CDCClass I Obese	0.376* (0.225)	0.214* (0.122)
BMI_CDCClass II Obese	0.863** (0.348)	0.479*** (0.180)
BMI_CDCClass III Obese	0.065 (0.368)	0.037 (0.202)
BMI_CDCOverweight	0.225 (0.188)	0.117 (0.102)
BMI_CDCUnder_Weight	0.284 (0.499)	0.164 (0.274)
SmokerChain Smokers	-0.505* (0.297)	-0.298* (0.166)
SmokerOccasional	-0.121 (0.428)	-0.090 (0.228)
SmokerUndisclosed	-0.392 (0.256)	-0.216 (0.143)
Age	-0.023*** (0.005)	-0.013*** (0.003)
GenderMale	-0.761 (0.714)	-0.418 (0.387)
HasPartnerYes	0.193 (0.163)	0.097 (0.088)
RaceAfrican American	0.056 (0.378)	0.075 (0.209)
RaceAsian	0.205 (0.406)	0.169 (0.223)
RaceWhite	0.309 (0.348)	0.227 (0.192)
IncomeRangesMiddle	0.200 (0.214)	0.107 (0.115)
IncomeRangesUpper	0.260 (0.202)	0.126 (0.106)
MedConditions_CatAnxiety	0.678** (0.336)	0.361** (0.169)
MedConditions_CatBP	0.173 (0.212)	0.102 (0.114)
MedConditions_CatDiabetes	-0.155 (0.232)	-0.072 (0.128)
MedConditions_CatHeart Condition	-0.468 (0.455)	-0.245 (0.261)
MedConditions_CatLung Disease	0.549 (0.343)	0.298* (0.177)
Health_CatFair	0.117 (0.171)	0.039 (0.093)
Health_CatPoor	2.298** (1.043)	1.125** (0.460)
LifestyleModerate	-0.309 (0.306)	-0.150 (0.156)
LifestyleSedentary	-0.867*** (0.322)	-0.444*** (0.167)
PandemicYes	-0.080 (0.160)	-0.037 (0.086)
AlcoholIntakeHigh	0.051 (1.016)	-0.022 (0.518)
AlcoholIntakeLow	-0.433 (0.718)	-0.240 (0.390)
AlcoholIntakeModerate	-0.380 (0.780)	-0.172 (0.421)
Constant	3.345*** (1.003)	1.871*** (0.547)
Observations	1,690	1,690
Log Likelihood	-594.915	-595.054
Akaike Inf. Crit.	1,259.829	1,260.109

Interpretation:

Explaining the interpretation of the logit and probit model using the Health Insurance variable

Probit - if the health insurance variable changes from 0 to 1 that is if the person gets health insurance, then the z-score of the probability of P(Y=1) that is a person using the health wellness app increase by 0.167.
Logit - if the health insurance variable changes from 0 to 1 that is if the person gets health insurance, then the expected change in log(odds) is 0.352.

With the Probit model we have certain better precision since the probability distribution function in Probit is Normal distribution whereas the probability distribution function in Logit is Logistic distribution, it is closely similar to normal distribution but not exactly normal, and thus with logit models we may have non-constant error variances, but estimates of Logit model are easier to interpret in terms of odds ratio (odds ratio = P/1-P i.e. probability of success/ probability of failure) whereas with Probit model explaining the estimates in terms of z-score is hard to practically understand and hence we decided to go for Logit to further explain consumer perceptions.

Evaluating the Model

```
> Confmatp
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0   6   1
         1  55 371

               Accuracy : 0.8707
                 95% CI : (0.8354, 0.9008)
    No Information Rate : 0.8591
    P-Value [Acc > NIR] : 0.2705

                  Kappa : 0.1519

 Mcnemar's Test P-Value : 1.417e-12

            Sensitivity : 0.09836
            Specificity : 0.99731
         Pos Pred Value : 0.85714
         Neg Pred Value : 0.87089
             Prevalence : 0.14088
         Detection Rate : 0.01386
   Detection Prevalence : 0.01617
      Balanced Accuracy : 0.54784

       'Positive' Class : 0
```

Since we are doing classification analysis, we have divided our data set into test and train data set where we have used 75% of the data for training and 25% of the data for testing.

Confusion Matrix:

		Prediction	
		0	1
Actual	0	6 *True Negative*	1 *False Positive*
	1	55 *False Negative*	371 *True Positive*

Statistics:

- ❖ Accuracy = (TP+TN)/(TP+TN+FP+FN) = 87%
- ❖ Error Rate = (FP+FN)/(TP+TN+FP+FN) = 13%
- ❖ Specificity = TN / (TN+FP) = 0.99
- ❖ Precision = TP / (TP+FP) = 0.86
- ❖ Recall = TP / (TP+FN) = 0.09
- ❖ F1-score = 2*Recall*Precision / (Recall + Precision) = 0.17

Model Insights and Output

	Coef	Odd	Prb
(Intercept)	3.3454	28.37193	0.965954
Health_CatPoor	2.298417	9.958407	**0.908746**
BMI_CDCClass II Obese	0.862735	2.369633	**0.703232**
MedConditions_CatAnxiety	0.678472	1.970865	**0.663398**
MedConditions_CatLung Disease	0.549237	1.731932	**0.633959**
BMI_CDCClass I Obese	0.37556	1.455806	0.592802
Health.InsuranceYes	0.352253	1.422268	**0.587164**
RaceWhite	0.30919	1.362321	0.576687
BMI_CDCUnder_Weight	0.284178	1.32867	0.57057
IncomeRangesUpper	0.260299	1.297317	**0.56471**
BMI_CDCOverweight	0.224868	1.252158	0.555981
RaceAsian	0.205427	1.228049	0.551177
IncomeRangesMiddle	0.199857	1.221228	0.549799
CaregivingYes	0.195617	1.216061	0.548749
HasPartnerYes	0.193363	1.213324	0.548191
MedConditions_CatBP	0.172812	1.188642	0.543096
Health_CatFair	0.11665	1.123726	0.52913
BMI_CDCClass III Obese	0.064719	1.06686	0.516174

	Coef	Odd	Prb
RaceAfrican American	0.055986	1.057583	0.513993
AlcoholIntakeHigh	0.051264	1.0526	**0.512813**
Age	-0.02269	0.977565	**0.494328**
PandemicYes	-0.07977	0.923324	**0.480067**
SmokerOccasional	-0.12128	0.885788	0.469718
MedConditions_CatDiabetes	-0.15508	0.856343	0.461307
LifestyleModerate	-0.30869	0.734406	0.423434
AlcoholIntakeModerate	-0.37984	0.683972	0.406166
SmokerUndisclosed	-0.39227	0.67552	0.40317
AlcoholIntakeLow	-0.43287	0.648642	0.39344
MedConditions_CatHeart Condition	-0.46837	0.626021	0.385002
SmokerChain Smokers	-0.50493	0.603547	0.376383
have.deviceSmartphone	-0.73411	0.479933	0.324294
GenderMale	-0.76129	0.467062	0.318365
have.deviceTablet/Computer only	-0.79434	0.451879	0.311237
LifestyleSedentary	-0.86698	0.420218	0.295883
have.deviceNone	-2.76254	0.063131	0.059382

Some Important findings from the above table:

- Health Category – People who have poor health tend to use the app more than people who are healthy by a probability of 91%.
- Medical Conditions - Though people with anxiety and lung disease have a higher probability of using a health wellness app but people with a heart condition, diabetes and BP have a lower probability.
- BMI – People with BMI categorized as obese have a higher probability of using health wellness app by a probability of 70% than healthy people.
- Gender – Males have a lower probability of using health wellness app as compared to females.
- Alcohol Intake – People categorized with high alcohol intake have a higher probability of using health wellness app by a probability of 51% as compared to normal intake.
- Lifestyle – People categorized with lifestyle as active have a higher probability of using health wellness app as compared to people with a moderate and sedentary lifestyle.

❖ Pandemic – From the data we have we found that the people in Pandemic are using the health and wellness app less by a probability of 48% as compared to before pandemic and the reason for that is that the data we have is mostly from before Pandemic and the data during a pandemic is not a lot so we cannot say with confidence about the use of the health wellness app by people during Pandemic.

Further scope

In this project, we examined the issues from the perspective of individual users, as opposed to that of health providers, and explored the contextual condition that influences one's decisions to engage in or disengage from health-related use of digital technologies. But in this project, we have not considered the effect of the privacy factor. And as you know with the rapid digitalization of medical records Privacy is a critical ingredient to the success of any electronic-based health service. So as a further investigation We can explore whether and to what extent privacy concerns and confidence regarding medical data may discourage or encourage people's engagement in health wellness app.

References

- https://www.healthitoutcomes.com/doc/mhealth-statistics-to-know-0001
- https://news.gallup.com/poll/269096/one-five-adults-health-apps-wearable-trackers.aspx
- https://www.ortholive.com/blog/32-statistics-on-mhealth/
- Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review - PubMed (nih.gov)
- Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics - PubMed (nih.gov)
- Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach - PubMed (nih.gov)