Project1-Group2

Emma Zhang, Ricky Wong, Xiaoxin Bian

2024-06-18

## Objective

The goal of this project is to conduct a sentiment analysis of popular keywords and phrases across Twitter, Instagram, and Facebook. This analysis aims to uncover and understand public opinions and emotions, providing actionable insights that enable Ignite Digital, a digital marketing agency, to refine their strategies and enhance their service offerings.

## Install required libraries

## Load the dataset

df <- read.csv("./sentimentdataset.csv")

**Use glimpse to explore the dataset.**

glimpse(df)

## Rows: 732  
## Columns: 15  
## $ X <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1…  
## $ Unnamed..0 <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1…  
## $ Text <chr> " Enjoying a beautiful day at the park! ", " T…  
## $ Sentiment <chr> " Positive ", " Negative ", " Positive ", " Positive ",…  
## $ Timestamp <chr> "2023-01-15 12:30:00", "2023-01-15 08:45:00", "2023-01-15 1…  
## $ User <chr> " User123 ", " CommuterX ", " FitnessFan ", " Adv…  
## $ Platform <chr> " Twitter ", " Twitter ", " Instagram ", " Facebook ", " …  
## $ Hashtags <chr> " #Nature #Park ", " #Traffic #M…  
## $ Retweets <dbl> 15, 5, 20, 8, 12, 25, 10, 15, 30, 18, 22, 7, 12, 28, 15, 20…  
## $ Likes <dbl> 30, 10, 40, 15, 25, 50, 20, 30, 60, 35, 45, 15, 25, 55, 30,…  
## $ Country <chr> " USA ", " Canada ", " USA ", " UK ", "…  
## $ Year <int> 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023,…  
## $ Month <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ Day <int> 15, 15, 15, 15, 15, 16, 16, 16, 17, 17, 17, 18, 18, 18, 19,…  
## $ Hour <int> 12, 8, 15, 18, 19, 9, 14, 19, 8, 12, 15, 10, 14, 18, 9, 13,…

## Data Processing

**Remove unnecessary columns**

df <- df[, -c(1, 2)]  
df$Country <- as.factor(str\_trim(df$Country))

**Define sentiment label categories**

positive <- c("Positive", "Happiness", "Joy", "Love", "Amusement", "Enjoyment", "Admiration", "Affection", "Awe","Acceptance", "Adoration", "Calmness", "Excitement", "Kind", "Pride", "Elation", "Euphoria", "Contentment", "Serenity", "Gratitude", "Hope", "Empowerment", "Compassion", "Tenderness", "Enthusiasm", "Fulfillment", "Reverence", "Hopeful", "Proud", "Grateful", "Empathetic", "Compassionate", "Playful", "Free-spirited", "Inspired", "Confident", "Thrill", "Overjoyed", "Inspiration", "Motivation", "JoyfulReunion","Satisfaction", "Blessed", "Appreciation", "Confidence", "Accomplishment", "Wonderment", "Optimism", "Enchantment", "Mindfulness", "Elegance", "Whimsy", "Harmony", "Creativity", "Radiance", "Wonder", "Rejuvenation", "Coziness", "Adventure", "Melodic", "FestiveJoy", "InnerJourney", "Freedom", "Dazzle", "Adrenaline", "ArtisticBurst", "CulinaryOdyssey", "Resilience", "Immersion", "Spark", "Marvel", "Positivity", "Kindness", "Friendship", "Success", "Exploration", "Amazement", "Romance", "Captivation", "Tranquility", "Grandeur", "Emotion", "Energy", "Celebration", "Charm", "Ecstasy", "Colorful", "Hypnotic", "Connection", "Iconic", "Journey", "Engagement", "Touched", "Triumph", "Heartwarming", "Solace", "Breakthrough", "Imagination", "Vibrancy", "Mesmerizing", "Creative Inspiration", "Nature's Beauty", "Celestial Wonder", "Happy", "PlayfulJoy", "DreamChaser", "Blessing", "Sympathy", "Renewed Effort", "Culinary Adventure", "Determination", "Zest", "Joy in Baking", "Challenge", "Winter Magic", "Thrilling Journey", "Runway Creativity", "Ocean's Freedom", "Relief", "Curiosity"  
)  
  
negative <- c("Negative", "Anger", "Fear", "Sadness", "Disgust", "Disappointed", "Bitter", "Shame", "Despair", "Grief", "Loneliness", "Jealousy", "Resentment", "Frustration", "Boredom", "Anxiety", "Intimidation", "Helplessness", "Envy", "Regret", "Melancholy", "Fearful", "Apprehensive", "Overwhelmed", "Devastated", "Frustrated", "Envious", "Dismissive", "Bitterness", "Heartbreak", "Betrayal", "Suffering", "EmotionalStorm", "Isolation", "Disappointment", "LostLove", "Exhaustion", "Sorrow", "Darkness", "Desperation", "Ruins", "Desolation", "Loss", "Heartache", "Solitude", "Obstacle", "Pressure", "Miscalculation", "Sad", "Hate", "Bad", "Jealous", "Embarrassed"  
)  
  
neutral <- c("Neutral", "Surprise", "Anticipation", "Confusion", "Arousal", "Indifference", "Numbness", "Nostalgia", "Ambivalence", "Contemplation", "Reflection", "Intrigue", "Whispers of the Past", "Pensive", "Bittersweet", "Suspense", "Envisioning History", "Mischievous", "Yearning"  
)

**Add a label column based on category definition**

df <- df %>% mutate(  
 Sentiment = trimws(Sentiment),  
 Sentiment\_Label = case\_when(  
 Sentiment %in% positive ~ "Positive",  
 Sentiment %in% negative ~ "Negative",  
 Sentiment %in% neutral ~ "Neutral"  
))

**View processed data**

head(df)

## Text Sentiment  
## 1 Enjoying a beautiful day at the park! Positive  
## 2 Traffic was terrible this morning. Negative  
## 3 Just finished an amazing workout! 💪 Positive  
## 4 Excited about the upcoming weekend getaway! Positive  
## 5 Trying out a new recipe for dinner tonight. Neutral  
## 6 Feeling grateful for the little things in life. Positive  
## Timestamp User Platform  
## 1 2023-01-15 12:30:00 User123 Twitter   
## 2 2023-01-15 08:45:00 CommuterX Twitter   
## 3 2023-01-15 15:45:00 FitnessFan Instagram   
## 4 2023-01-15 18:20:00 AdventureX Facebook   
## 5 2023-01-15 19:55:00 ChefCook Instagram   
## 6 2023-01-16 09:10:00 GratitudeNow Twitter   
## Hashtags Retweets Likes Country Year  
## 1 #Nature #Park 15 30 USA 2023  
## 2 #Traffic #Morning 5 10 Canada 2023  
## 3 #Fitness #Workout 20 40 USA 2023  
## 4 #Travel #Adventure 8 15 UK 2023  
## 5 #Cooking #Food 12 25 Australia 2023  
## 6 #Gratitude #PositiveVibes 25 50 India 2023  
## Month Day Hour Sentiment\_Label  
## 1 1 15 12 Positive  
## 2 1 15 8 Negative  
## 3 1 15 15 Positive  
## 4 1 15 18 Positive  
## 5 1 15 19 Neutral  
## 6 1 16 9 Positive

**Sentiment distribution by platform**

sentiment\_platform <- df %>%  
 mutate(Platform = trimws(Platform)) %>%   
 group\_by(Platform, Sentiment\_Label) %>%  
 summarise(Count = n())

## `summarise()` has grouped output by 'Platform'. You can override using the  
## `.groups` argument.

sentiment\_platform

## # A tibble: 9 × 3  
## # Groups: Platform [3]  
## Platform Sentiment\_Label Count  
## <chr> <chr> <int>  
## 1 Facebook Negative 56  
## 2 Facebook Neutral 22  
## 3 Facebook Positive 153  
## 4 Instagram Negative 60  
## 5 Instagram Neutral 34  
## 6 Instagram Positive 164  
## 7 Twitter Negative 64  
## 8 Twitter Neutral 28  
## 9 Twitter Positive 151

extract\_hashtags <- function(hashtags) {  
 unlist(strsplit(hashtags, " "))  
}  
  
all\_hashtags <- unlist(lapply(df$Hashtags, extract\_hashtags))  
all\_hashtags <- all\_hashtags[all\_hashtags != "" & all\_hashtags != " "]  
hashtag\_counts <- as.data.frame(table(all\_hashtags))  
  
colnames(hashtag\_counts) <- c("Hashtag", "Frequency")  
sorted\_hashtag\_counts <- hashtag\_counts %>%  
 arrange(desc(Frequency))  
  
sorted\_hashtag\_counts

## Hashtag Frequency  
## 1 #Serenity 15  
## 2 #Excitement 13  
## 3 #Gratitude 13  
## 4 #Despair 11  
## 5 #Nostalgia 11  
## 6 #Contentment 10  
## 7 #Curiosity 10  
## 8 #Awe 9  
## 9 #Grief 9  
## 10 #Hopeful 9  
## 11 #Loneliness 9  
## 12 #Acceptance 8  
## 13 #Confusion 8  
## 14 #Enthusiasm 8  
## 15 #Joy 8  
## 16 #Determination 7  
## 17 #Elation 7  
## 18 #Euphoria 7  
## 19 #Inspiration 7  
## 20 #Ambivalence 6  
## 21 #Bitterness 6  
## 22 #Melancholy 6  
## 23 #Numbness 6  
## 24 #Playful 6  
## 25 #Pride 6  
## 26 #Reflection 6  
## 27 #Regret 6  
## 28 #Adventure 5  
## 29 #Betrayal 5  
## 30 #Empowerment 5  
## 31 #Frustrated 5  
## 32 #Happiness 5  
## 33 #Hope 5  
## 34 #Indifference 5  
## 35 #Inspired 5  
## 36 #Overwhelmed 5  
## 37 #Admiration 4  
## 38 #Arousal 4  
## 39 #Calmness 4  
## 40 #Compassion 4  
## 41 #Compassionate 4  
## 42 #Creativity 4  
## 43 #Desolation 4  
## 44 #Disgust 4  
## 45 #Enchantment 4  
## 46 #Frustration 4  
## 47 #Fulfillment 4  
## 48 #Grateful 4  
## 49 #HighSchoolArt 4  
## 50 #Injustice 4  
## 51 #Kindness 4  
## 52 #Love 4  
## 53 #Nature 4  
## 54 #Proud 4  
## 55 #Resentment 4  
## 56 #Reverence 4  
## 57 #Surprise 4  
## 58 #Tenderness 4  
## 59 #Thrill 4  
## 60 #Accomplishment 3  
## 61 #Amusement 3  
## 62 #Anticipation 3  
## 63 #BookTime 3  
## 64 #Boredom 3  
## 65 #CarnivalEscapade 3  
## 66 #Celebration 3  
## 67 #Confident 3  
## 68 #Cooking 3  
## 69 #Devastated 3  
## 70 #Disappointment 3  
## 71 #Dismissive 3  
## 72 #Empathetic 3  
## 73 #Envious 3  
## 74 #Envy 3  
## 75 #FamilyLaughter 3  
## 76 #Fear 3  
## 77 #Fearful 3  
## 78 #FitnessGoals 3  
## 79 #FreeSpirit 3  
## 80 #Harmony 3  
## 81 #Heartbreak 3  
## 82 #HighSchoolAthlete 3  
## 83 #HighSchoolAwkward 3  
## 84 #HighSchoolMemories 3  
## 85 #HighSchoolStruggles 3  
## 86 #Isolation 3  
## 87 #Jealous 3  
## 88 #Jealousy 3  
## 89 #Loss 3  
## 90 #Milestone 3  
## 91 #Mindfulness 3  
## 92 #Optimism 3  
## 93 #PersonalGrowth 3  
## 94 #ScalingPeaks 3  
## 95 #SeedsOfOptimism 3  
## 96 #Shame 3  
## 97 #Success 3  
## 98 #TearsOfEmpathy 3  
## 99 #TeenLife 3  
## 100 #Wonder 3  
## 101 #Adoration 2  
## 102 #Affection 2  
## 103 #AncientMysteries 2  
## 104 #Anxiety 2  
## 105 #Apprehensive 2  
## 106 #ArchitecturalGrandeur 2  
## 107 #Art 2  
## 108 #Blessed 2  
## 109 #BrighterTomorrows 2  
## 110 #Captivation 2  
## 111 #Challenges 2  
## 112 #CloudsOfCreativity 2  
## 113 #Contemplation 2  
## 114 #Corruption 2  
## 115 #Coziness 2  
## 116 #CulturalCelebration 2  
## 117 #DanceOfLife 2  
## 118 #Disappointed 2  
## 119 #DriftingThroughLife 2  
## 120 #Emotion 2  
## 121 #EmotionalOutburst 2  
## 122 #EmotionalShield 2  
## 123 #EmotionalStorm 2  
## 124 #EmotionalStorms 2  
## 125 #Enjoyment 2  
## 126 #Exploration 2  
## 127 #FamilyTime 2  
## 128 #Farewell 2  
## 129 #Fitness 2  
## 130 #ForestWhispers 2  
## 131 #Friendship 2  
## 132 #FriendshipGoals 2  
## 133 #Gaming 2  
## 134 #GoldenHues 2  
## 135 #Helplessness 2  
## 136 #HiddenGem 2  
## 137 #HighSchoolCelebration 2  
## 138 #HighSchoolCreativity 2  
## 139 #HighSchoolEmbarrassment 2  
## 140 #HighSchoolEmotions 2  
## 141 #HighSchoolProjects 2  
## 142 #HighSchoolReality 2  
## 143 #HighSchoolRegret 2  
## 144 #HighSchoolScience 2  
## 145 #HighSchoolSkills 2  
## 146 #HighSchoolSnacks 2  
## 147 #HighSchoolSports 2  
## 148 #HighSchoolTech 2  
## 149 #IndifferencePerformance 2  
## 150 #InnerPeace 2  
## 151 #Intimidation 2  
## 152 #JugglingResponsibilities 2  
## 153 #LanternOfCompassion 2  
## 154 #LifeJourney 2  
## 155 #LifeLessons 2  
## 156 #LostLove 2  
## 157 #Meditation 2  
## 158 #Memories 2  
## 159 #MissedChances 2  
## 160 #MosaicOfDisappointment 2  
## 161 #MovieNight 2  
## 162 #MusicalUnity 2  
## 163 #NewBeginnings 2  
## 164 #OldFriends 2  
## 165 #PetLove 2  
## 166 #PositiveVibes 2  
## 167 #Radiance 2  
## 168 #RainyDays 2  
## 169 #Reading 2  
## 170 #Rejuvenation 2  
## 171 #Resilience 2  
## 172 #Sadness 2  
## 173 #SailorOfDreams 2  
## 174 #Satisfaction 2  
## 175 #SeasideEscape 2  
## 176 #SeniorCreativity 2  
## 177 #SeniorJazz 2  
## 178 #SeniorTravel 2  
## 179 #ShatteredTrust 2  
## 180 #SkillBuilding 2  
## 181 #SolitaryNocturne 2  
## 182 #Sorrow 2  
## 183 #SpreadLove 2  
## 184 #SpringBlooms 2  
## 185 #StandUpComedy 2  
## 186 #StargazingAdventure 2  
## 187 #StarryNight 2  
## 188 #SupportiveCommunity 2  
## 189 #SurpriseCelebration 2  
## 190 #SurpriseParty 2  
## 191 #TalesToTheStars 2  
## 192 #ThoughtLabyrinth 2  
## 193 #TimelessTunes 2  
## 194 #Tranquility 2  
## 195 #TranquilWaters 2  
## 196 #Travel 2  
## 197 #VirtualReality 2  
## 198 #WarmWinterEvening 2  
## 199 #Whimsy 2  
## 200 #WingsOfDreams 2  
## 201 #Yearning 2  
## 202 #Zest 2  
## 203 #AbstractArt 1  
## 204 #AbyssOfHeartache 1  
## 205 #AcademicFrustration 1  
## 206 #Accomplished 1  
## 207 #Accomplishments 1  
## 208 #AchieveTheImpossible 1  
## 209 #ActsOfKindness 1  
## 210 #AdeleConcert 1  
## 211 #AdeleMelodies 1  
## 212 #Adrenaline 1  
## 213 #AdventureAwaits 1  
## 214 #AdventureTime 1  
## 215 #AgedToPerfection 1  
## 216 #AgingGracefully 1  
## 217 #AI 1  
## 218 #Amazement 1  
## 219 #AncientCivilization 1  
## 220 #AncientCivilizations 1  
## 221 #AncientDiscovery 1  
## 222 #AncientTales 1  
## 223 #Anger 1  
## 224 #Angry 1  
## 225 #Appreciation 1  
## 226 #Argument 1  
## 227 #ArianaGrande 1  
## 228 #ArtAppreciation 1  
## 229 #ArtCollaboration 1  
## 230 #ArtDiscovery 1  
## 231 #ArtisticBurst 1  
## 232 #ArtisticExpression 1  
## 233 #ArtisticMishaps 1  
## 234 #ArtisticProject 1  
## 235 #ArtistsHaven 1  
## 236 #AthleticAchievement 1  
## 237 #AutumnLeaves 1  
## 238 #AutumnWaltz 1  
## 239 #AvoidingEyeContact 1  
## 240 #BadDay 1  
## 241 #BadHairDay 1  
## 242 #BadLuck 1  
## 243 #BadLuckDay 1  
## 244 #BadService 1  
## 245 #BakingAdventures 1  
## 246 #BalancingAct 1  
## 247 #BallroomDance 1  
## 248 #BasketballFinals 1  
## 249 #BasketballStruggles 1  
## 250 #BeachDays 1  
## 251 #BeautifulChaos 1  
## 252 #BeyonceConcert 1  
## 253 #BikeRide 1  
## 254 #BingeWatchDay 1  
## 255 #Birthday 1  
## 256 #BirthdayParty 1  
## 257 #BirthdaySurprise 1  
## 258 #Bitter 1  
## 259 #BitterAftertaste 1  
## 260 #BitterExperience 1  
## 261 #BitterLesson 1  
## 262 #BitterPill 1  
## 263 #Bittersweet 1  
## 264 #BittersweetMemories 1  
## 265 #Blogging 1  
## 266 #BloggingJourney 1  
## 267 #BloomingAchievements 1  
## 268 #BlossomingConnection 1  
## 269 #BobMarleyTribute 1  
## 270 #BookClub 1  
## 271 #BookDiscoveries 1  
## 272 #BookishEscape 1  
## 273 #BookishTalks 1  
## 274 #BookRelease 1  
## 275 #BouldersOfExhaustion 1  
## 276 #BoundlessLaughter 1  
## 277 #BoxingDefeat 1  
## 278 #BranchesOfAchievement 1  
## 279 #Breakthrough 1  
## 280 #Breathtaking 1  
## 281 #BreathtakingSunrise 1  
## 282 #BrightFuture 1  
## 283 #BrilliantTrails 1  
## 284 #BrokenTrust 1  
## 285 #Brunch 1  
## 286 #BrunoMars 1  
## 287 #BurgerQuest 1  
## 288 #BusinessVenture 1  
## 289 #CafeReflection 1  
## 290 #CafeteriaDrama 1  
## 291 #CalmAfterStorm 1  
## 292 #CamaraderieBlooms 1  
## 293 #CanvasOfDreams 1  
## 294 #CaptivatingNovel 1  
## 295 #Career 1  
## 296 #CareerDevelopment 1  
## 297 #CareerMilestone 1  
## 298 #CarefreeSpirit 1  
## 299 #CartwheelChallenge 1  
## 300 #CelestialWonder 1  
## 301 #CelestialWonders 1  
## 302 #Challenge 1  
## 303 #ChallengeAccepted 1  
## 304 #CharityEvent 1  
## 305 #Charm 1  
## 306 #ChemistryWizard 1  
## 307 #CherishedMemories 1  
## 308 #CherishedMoment 1  
## 309 #CherishedPlaces 1  
## 310 #ChildhoodFantasies 1  
## 311 #ChildhoodFavorites 1  
## 312 #ChildhoodMemories 1  
## 313 #ChildrensPlaydate 1  
## 314 #CityExplore 1  
## 315 #ClassEvent 1  
## 316 #ClassicFilmMoments 1  
## 317 #ClosingWalls 1  
## 318 #CoastalAdventure 1  
## 319 #Coding 1  
## 320 #CodingAdventure 1  
## 321 #CodingTriumph 1  
## 322 #CoffeeJourney 1  
## 323 #CoffeeTime 1  
## 324 #ColdplayMagic 1  
## 325 #CollegeDreams 1  
## 326 #Colorful 1  
## 327 #ColorfulStreetMarket 1  
## 328 #ColorfulSunset 1  
## 329 #ComedyShow 1  
## 330 #Community 1  
## 331 #CommunityArt 1  
## 332 #CommunityGiving 1  
## 333 #CommunityKindness 1  
## 334 #CommunityService 1  
## 335 #CommunitySupport 1  
## 336 #ConcertNight 1  
## 337 #Confidence 1  
## 338 #ConflictingEmotions 1  
## 339 #ConflictingFeelings 1  
## 340 #Connection 1  
## 341 #CookingFail 1  
## 342 #CookingSuccess 1  
## 343 #CosmicFireflies 1  
## 344 #Cozy 1  
## 345 #CozyEvening 1  
## 346 #CozyNight 1  
## 347 #CrashingWaves 1  
## 348 #CreativeExpressions 1  
## 349 #CreativeInspiration 1  
## 350 #CreativePhoenix 1  
## 351 #CreativityAtAnyAge 1  
## 352 #CricketChampionship 1  
## 353 #CrimeSeriesThrills 1  
## 354 #CrossroadsConfusion 1  
## 355 #CrushedHopes 1  
## 356 #CrushedSpirit 1  
## 357 #CrushFail 1  
## 358 #CrushStruggles 1  
## 359 #CulinaryAdventure 1  
## 360 #CulinaryJourney 1  
## 361 #CulinaryOdyssey 1  
## 362 #CulturalDay 1  
## 363 #CulturalExperience 1  
## 364 #CustomerService 1  
## 365 #CyclingChampion 1  
## 366 #CyclingFrustration 1  
## 367 #DanceFail 1  
## 368 #DanceInTheRain 1  
## 369 #DanceLife 1  
## 370 #DarkDays 1  
## 371 #Darkness 1  
## 372 #Dazzle 1  
## 373 #Debate 1  
## 374 #DebateSkills 1  
## 375 #DebateWarrior 1  
## 376 #DecisionMaking 1  
## 377 #DeepDescent 1  
## 378 #DeepMeanings 1  
## 379 #Desire 1  
## 380 #DesireForKeys 1  
## 381 #Desperation 1  
## 382 #Dessert 1  
## 383 #DiaryJourney 1  
## 384 #DigitalOverload 1  
## 385 #DiscoverNewPlaces 1  
## 386 #DistantWarmth 1  
## 387 #Diversity 1  
## 388 #DIYGoals 1  
## 389 #DIYProject 1  
## 390 #DoodleMaster 1  
## 391 #DoubleRainbow 1  
## 392 #DramaClubStar 1  
## 393 #DreamChaser 1  
## 394 #DreamsIntoReality 1  
## 395 #DrowningSoul 1  
## 396 #EarthsGrandeur 1  
## 397 #EchoesOfShatteredDreams 1  
## 398 #EchoingSilence 1  
## 399 #Ecstasy 1  
## 400 #EdSheeranConcert 1  
## 401 #EfficiencyGoals 1  
## 402 #Elegance 1  
## 403 #EmotionalWinter 1  
## 404 #EndlessDarkness 1  
## 405 #EndlessReverberation 1  
## 406 #EndlessSmiles 1  
## 407 #Energy 1  
## 408 #Engagement 1  
## 409 #EngineeringMarvels 1  
## 410 #Environment 1  
## 411 #EnvironmentalEfforts 1  
## 412 #EnvisioningHistory 1  
## 413 #Exhaustion 1  
## 414 #ExtraordinaryPath 1  
## 415 #FacingFear 1  
## 416 #FadingPossibilities 1  
## 417 #FairyTaleMagic 1  
## 418 #FamilyDramaMoments 1  
## 419 #FamilyLove 1  
## 420 #FamilyPhotoAlbum 1  
## 421 #FantasyBookNerd 1  
## 422 #FashionGoals 1  
## 423 #FashionInnovation 1  
## 424 #FavoriteAuthor 1  
## 425 #FearlessExplorer 1  
## 426 #FestivalExcitement 1  
## 427 #FestiveJoy 1  
## 428 #FieldsOfEnthusiasm 1  
## 429 #FireflyField 1  
## 430 #FirstDayOops 1  
## 431 #FitnessChallenge 1  
## 432 #FitnessChallengeCompleted 1  
## 433 #FitnessMilestone 1  
## 434 #FitnessSuccess 1  
## 435 #FlavorsAroundTheWorld 1  
## 436 #FloatingThroughLife 1  
## 437 #FloralJoy 1  
## 438 #Food 1  
## 439 #FootprintsOfAssuredness 1  
## 440 #ForgottenEra 1  
## 441 #ForgottenLove 1  
## 442 #Formula1Championship 1  
## 443 #Freedom 1  
## 444 #FreshBreadAroma 1  
## 445 #Friends 1  
## 446 #FriendshipAdventures 1  
## 447 #FriendshipBracelets 1  
## 448 #FriendshipReunion 1  
## 449 #FriendshipSunshine 1  
## 450 #FulfillingEnding 1  
## 451 #Fun 1  
## 452 #FunTimes 1  
## 453 #FurryFriend 1  
## 454 #Future 1  
## 455 #GalleryExplore 1  
## 456 #GameOn 1  
## 457 #GardenFriends 1  
## 458 #Gardening 1  
## 459 #GetawayGoals 1  
## 460 #Gift 1  
## 461 #GildedPrize 1  
## 462 #GladiatorTales 1  
## 463 #GlimmerOfHope 1  
## 464 #GlobalFriendship 1  
## 465 #GNRConcert 1  
## 466 #GoldenGlow 1  
## 467 #GolfDefeat 1  
## 468 #GolfVictory 1  
## 469 #GracefulMoves 1  
## 470 #Grandeur 1  
## 471 #GreenEyedMonster 1  
## 472 #Growth 1  
## 473 #GuidingStar 1  
## 474 #GymnasticsFall 1  
## 475 #HackySackChallenge 1  
## 476 #HairExperiment 1  
## 477 #HandcraftedArt 1  
## 478 #HandstandChallenge 1  
## 479 #HappinessInFood 1  
## 480 #HappinessMilestone 1  
## 481 #HauntedPast 1  
## 482 #Health 1  
## 483 #HealthyLiving 1  
## 484 #Heartache 1  
## 485 #HeartacheSymphony 1  
## 486 #HeartOfDiscontent 1  
## 487 #Heartwarming 1  
## 488 #HeartwarmingMoments 1  
## 489 #HeavyHeart 1  
## 490 #HelpingOthers 1  
## 491 #HiddenGemDiscovery 1  
## 492 #HiddenGems 1  
## 493 #HighSchoolActor 1  
## 494 #HighSchoolAdventures 1  
## 495 #HighSchoolAthletics 1  
## 496 #HighSchoolBakes 1  
## 497 #HighSchoolBlues 1  
## 498 #HighSchoolCoderLife 1  
## 499 #HighSchoolCommunity 1  
## 500 #HighSchoolConfidence 1  
## 501 #HighSchoolConnectivity 1  
## 502 #HighSchoolCorrespondence 1  
## 503 #HighSchoolCrafts 1  
## 504 #HighSchoolCriticism 1  
## 505 #HighSchoolDebater 1  
## 506 #HighSchoolDrama 1  
## 507 #HighSchoolEntertainment 1  
## 508 #HighSchoolExams 1  
## 509 #HighSchoolExperiments 1  
## 510 #HighSchoolExplorer 1  
## 511 #HighSchoolFairytale 1  
## 512 #HighSchoolFoodie 1  
## 513 #HighSchoolFriends 1  
## 514 #HighSchoolFriendship 1  
## 515 #HighSchoolFun 1  
## 516 #HighSchoolGamer 1  
## 517 #HighSchoolHardships 1  
## 518 #HighSchoolHobbies 1  
## 519 #HighSchoolIntellect 1  
## 520 #HighSchoolInterests 1  
## 521 #HighSchoolJobs 1  
## 522 #HighSchoolJourney 1  
## 523 #HighSchoolJoy 1  
## 524 #HighSchoolMagic 1  
## 525 #HighSchoolMusic 1  
## 526 #HighSchoolNostalgia 1  
## 527 #HighSchoolPhilanthropy 1  
## 528 #HighSchoolPhotographer 1  
## 529 #HighSchoolPositivity 1  
## 530 #HighSchoolReader 1  
## 531 #HighSchoolReads 1  
## 532 #HighSchoolScientist 1  
## 533 #HighSchoolSkater 1  
## 534 #HighSchoolStress 1  
## 535 #HighSchoolStyle 1  
## 536 #HighSchoolTalent 1  
## 537 #HighSchoolTension 1  
## 538 #HighSchoolTikTok 1  
## 539 #HighSchoolUnity 1  
## 540 #HighSchoolVolunteer 1  
## 541 #HighSchoolWeather 1  
## 542 #HighSchoolWellness 1  
## 543 #HikingAdventure 1  
## 544 #HistoricalSite 1  
## 545 #HistoricalTour 1  
## 546 #History 1  
## 547 #HockeyFinalsComeback 1  
## 548 #HolidaySupport 1  
## 549 #HomeCooking 1  
## 550 #HomeImprovement 1  
## 551 #HomemadeDelights 1  
## 552 #HopefulJourney 1  
## 553 #Hopeless 1  
## 554 #HuesOfSunset 1  
## 555 #Humanity 1  
## 556 #HumanityRestored 1  
## 557 #Hypnotic 1  
## 558 #IciclesOfResentment 1  
## 559 #Iconic 1  
## 560 #IcyMasterpiece 1  
## 561 #IgnitedSoul 1  
## 562 #Imagination 1  
## 563 #ImaginedHorrors 1  
## 564 #Immersion 1  
## 565 #ImpenetrableWalls 1  
## 566 #ImperfectJourney 1  
## 567 #IndieFilmSuccess 1  
## 568 #InnerJourney 1  
## 569 #Innovation 1  
## 570 #Insecurity 1  
## 571 #InspirationIgnited 1  
## 572 #InstrumentLearning 1  
## 573 #Integrity 1  
## 574 #Intrigue 1  
## 575 #JapaneseTraditions 1  
## 576 #JayZAnthem 1  
## 577 #JournalMoments 1  
## 578 #Journey 1  
## 579 #JoyfulReunion 1  
## 580 #JoyInBaking 1  
## 581 #JustinBieber 1  
## 582 #KatyPerry 1  
## 583 #KhmerStories 1  
## 584 #KitchenMemories 1  
## 585 #KnowledgeQuest 1  
## 586 #KpopFangirl 1  
## 587 #LabyrinthOfHopelessness 1  
## 588 #LabyrinthOfMiscommunication 1  
## 589 #LabyrinthOfQuestions 1  
## 590 #LabyrinthOfThoughts 1  
## 591 #LackOfInterest 1  
## 592 #LadyGaga 1  
## 593 #LakeTranquility 1  
## 594 #LanguageLearning 1  
## 595 #LateNight 1  
## 596 #Learning 1  
## 597 #LifeChallenges 1  
## 598 #LifeChoices 1  
## 599 #LifeInPictures 1  
## 600 #LifePuzzle 1  
## 601 #LifetimeAdventure 1  
## 602 #LifetimeExperiences 1  
## 603 #LifeWellLived 1  
## 604 #LiteraryEscape 1  
## 605 #LiveMusic 1  
## 606 #LonelyNights 1  
## 607 #LostFriendship 1  
## 608 #LostHeadphones 1  
## 609 #LostJoyFootsteps 1  
## 610 #LostLoveLonging 1  
## 611 #LoveLetterFail 1  
## 612 #LoveNote 1  
## 613 #LoveSongLonging 1  
## 614 #LoveStitches 1  
## 615 #LurkingMonsters 1  
## 616 #MagicalWorld 1  
## 617 #MagicFail 1  
## 618 #MakingSense 1  
## 619 #MarathonRecords 1  
## 620 #MarketStories 1  
## 621 #Marvel 1  
## 622 #MasqueradeElegance 1  
## 623 #MeditationBliss 1  
## 624 #MelancholicMelody 1  
## 625 #Melodic 1  
## 626 #MemoryLane 1  
## 627 #MentalHealth 1  
## 628 #Mentorship 1  
## 629 #Mesmerizing 1  
## 630 #MetallicaConcert 1  
## 631 #MindEnrichment 1  
## 632 #MindfulLiving 1  
## 633 #MinotaurOfPressure 1  
## 634 #MiragesOfConnection 1  
## 635 #Miscalculation 1  
## 636 #MissedMoments 1  
## 637 #MJTribute 1  
## 638 #Monotony 1  
## 639 #Mood 1  
## 640 #MoonlitPeace 1  
## 641 #MoonlitSerenade 1  
## 642 #Morning 1  
## 643 #MorningCoffee 1  
## 644 #Motivation 1  
## 645 #MoviePremiereThrills 1  
## 646 #MovieRelease 1  
## 647 #MovieTime 1  
## 648 #MuseumTales 1  
## 649 #Music 1  
## 650 #MusicHarmony 1  
## 651 #MusicLover 1  
## 652 #MusicMagic 1  
## 653 #MusicTherapy 1  
## 654 #MustWatch 1  
## 655 #MysteryNovel 1  
## 656 #NaturalDisaster 1  
## 657 #NatureAdventures 1  
## 658 #NatureCapture 1  
## 659 #NatureExploration 1  
## 660 #NatureOrchestra 1  
## 661 #NatureRetreat 1  
## 662 #NaturesBeauty 1  
## 663 #NatureTrailBlazer 1  
## 664 #NatureWhispers 1  
## 665 #NatureWonders 1  
## 666 #NewFamilyMember 1  
## 667 #NewHorizons 1  
## 668 #NewPost 1  
## 669 #NewProject 1  
## 670 #NewWeek 1  
## 671 #NewYear 1  
## 672 #NightDancePerformance 1  
## 673 #NightSkyTriumph 1  
## 674 #Obstacle 1  
## 675 #Obstacles 1  
## 676 #OceanRetreat 1  
## 677 #OceansFreedom 1  
## 678 #OldMemories 1  
## 679 #OlympicAchievement 1  
## 680 #OnlineGaming 1  
## 681 #OnlineHate 1  
## 682 #OnlineOpinions 1  
## 683 #OpenRoadAdventures 1  
## 684 #OrchestraMelody 1  
## 685 #OrigamiMaster 1  
## 686 #OscarWinningMoment 1  
## 687 #OutdoorBeauty 1  
## 688 #OutdoorClassroom 1  
## 689 #OvercomingAnxiety 1  
## 690 #Overjoyed 1  
## 691 #OwnGoal 1  
## 692 #PainfulEchoes 1  
## 693 #PaintingInProgress 1  
## 694 #PaintingTheSky 1  
## 695 #PaperPlaneExpert 1  
## 696 #ParadiseWhispers 1  
## 697 #ParisianDreams 1  
## 698 #Park 1  
## 699 #PastChoices 1  
## 700 #PastCurrents 1  
## 701 #PathOfRemorse 1  
## 702 #PeacefulMelody 1  
## 703 #PeacefulMind 1  
## 704 #Pensive 1  
## 705 #PerfectPuzzle 1  
## 706 #PerplexedMind 1  
## 707 #PersonalAchievement 1  
## 708 #PersonalDevelopment 1  
## 709 #PetAdoption 1  
## 710 #PetAntics 1  
## 711 #PhotoAlbumStories 1  
## 712 #PhotographyClub 1  
## 713 #PhotographyJourney 1  
## 714 #PhysicsSurvivor 1  
## 715 #PlayfulJoy 1  
## 716 #PlayfulKittens 1  
## 717 #PoisonedTrust 1  
## 718 #Politics 1  
## 719 #Positivity 1  
## 720 #Pressure 1  
## 721 #Productivity 1  
## 722 #ProductLaunchSuccess 1  
## 723 #Project 1  
## 724 #ProjectMiracle 1  
## 725 #ProjectStruggles 1  
## 726 #PromDreams 1  
## 727 #PuppyLove 1  
## 728 #PuzzleChallenge 1  
## 729 #PuzzleCompletion 1  
## 730 #QueenTribute 1  
## 731 #QuietMoments 1  
## 732 #QuietSerenade 1  
## 733 #QuietTime 1  
## 734 #RainDance 1  
## 735 #RaindropMelody 1  
## 736 #RareBookDiscovery 1  
## 737 #RenewedEffort 1  
## 738 #ResentmentWeb 1  
## 739 #Restlessness 1  
## 740 #ReverberatingPromises 1  
## 741 #RhythmOfAffection 1  
## 742 #RippleOfDespair 1  
## 743 #RoadTrip 1  
## 744 #RoadTripDownMemoryLane 1  
## 745 #RoadTripMemories 1  
## 746 #RollercoasterAdventure 1  
## 747 #RollercoasterThrills 1  
## 748 #Romance 1  
## 749 #Ruins 1  
## 750 #RumorMills 1  
## 751 #RunForACause 1  
## 752 #RunwayCreativity 1  
## 753 #SadnessWhirlwind 1  
## 754 #SayingGoodbye 1  
## 755 #ScatteredHope 1  
## 756 #ScenicTrails 1  
## 757 #ScenicViews 1  
## 758 #SchoolChallenges 1  
## 759 #SchoolClubs 1  
## 760 #SchoolEvents 1  
## 761 #ScienceFair 1  
## 762 #ScienceFairWinner 1  
## 763 #ScienceGeek 1  
## 764 #ScientificDiscovery 1  
## 765 #SculptingDreams 1  
## 766 #SeekerOfKnowledge 1  
## 767 #SeekingConnection 1  
## 768 #SelfCare 1  
## 769 #SelfieStruggle 1  
## 770 #SelfLoveJourney 1  
## 771 #SeniorArt 1  
## 772 #SeniorArtist 1  
## 773 #SeniorAstronomy 1  
## 774 #SeniorBlog 1  
## 775 #SeniorBonding 1  
## 776 #SeniorBookworms 1  
## 777 #SeniorCalm 1  
## 778 #SeniorCars 1  
## 779 #SeniorCheers 1  
## 780 #SeniorClicks 1  
## 781 #SeniorConcert 1  
## 782 #SeniorCulinary 1  
## 783 #SeniorCuriosity 1  
## 784 #SeniorCycling 1  
## 785 #SeniorDance 1  
## 786 #SeniorExhibition 1  
## 787 #SeniorExplorer 1  
## 788 #SeniorFamily 1  
## 789 #SeniorGarden 1  
## 790 #SeniorGreenThumb 1  
## 791 #SeniorLearning 1  
## 792 #SeniorLife 1  
## 793 #SeniorMemoir 1  
## 794 #SeniorMemories 1  
## 795 #SeniorMoments 1  
## 796 #SeniorMoves 1  
## 797 #SeniorNostalgia 1  
## 798 #SeniorReflections 1  
## 799 #SeniorSupport 1  
## 800 #SeniorTheater 1  
## 801 #SeniorVoices 1  
## 802 #SeniorWine 1  
## 803 #ShadesOfUncertainty 1  
## 804 #ShadowConversation 1  
## 805 #ShakiraRhythms 1  
## 806 #ShatteredDreams 1  
## 807 #ShatteredHopes 1  
## 808 #ShatteredVows 1  
## 809 #ShowerOfEmpathy 1  
## 810 #ShowerOfThankfulness 1  
## 811 #SickDay 1  
## 812 #SilentObserver 1  
## 813 #SilentSolitude 1  
## 814 #SilentSurrender 1  
## 815 #SilentTears 1  
## 816 #SimpleMoments 1  
## 817 #SinatraTribute 1  
## 818 #SkaterLife 1  
## 819 #SkyHighDreams 1  
## 820 #SleepoverFun 1  
## 821 #SlippingThreads 1  
## 822 #SmallJoys 1  
## 823 #SnackCrisis 1  
## 824 #SnackNinja 1  
## 825 #SnackSmuggler 1  
## 826 #SnowfallMagic 1  
## 827 #SnowPeakDance 1  
## 828 #SnowyLandscapes 1  
## 829 #SoccerDefeats 1  
## 830 #SocialMediaBlunder 1  
## 831 #Solace 1  
## 832 #Solitude 1  
## 833 #SoulEngulfed 1  
## 834 #SoulfulWriting 1  
## 835 #SoulUpliftment 1  
## 836 #Spark 1  
## 837 #SpecialDinner 1  
## 838 #SpecterOfTheUnknown 1  
## 839 #SpiritLifter 1  
## 840 #SportsTriumph 1  
## 841 #Spring 1  
## 842 #SpringtimeColors 1  
## 843 #StagePassion 1  
## 844 #Stagnation 1  
## 845 #StainedTrust 1  
## 846 #StandingTall 1  
## 847 #Stargazing 1  
## 848 #StarlessSky 1  
## 849 #StoicFacade 1  
## 850 #StoicResilience 1  
## 851 #StopCyberbullying 1  
## 852 #StormOfDoubt 1  
## 853 #StreetArtVibes 1  
## 854 #StreetFoodDelights 1  
## 855 #StreetMusic 1  
## 856 #Struggle 1  
## 857 #Struggling 1  
## 858 #StudioCreativity 1  
## 859 #StudyingMemes 1  
## 860 #StudyStrategies 1  
## 861 #Suffering 1  
## 862 #Summer 1  
## 863 #SummerCarnival 1  
## 864 #SundayMood 1  
## 865 #SunnyDay 1  
## 866 #SunriseHues 1  
## 867 #SunriseOfPossibilities 1  
## 868 #Sunset 1  
## 869 #SunsetBeauty 1  
## 870 #SunsetHues 1  
## 871 #SunsetMoments 1  
## 872 #SunsetViews 1  
## 873 #Supportive 1  
## 874 #SurfingThrills 1  
## 875 #SurpriseAdventure 1  
## 876 #Suspense 1  
## 877 #SwimmingDisappointment 1  
## 878 #Sympathy 1  
## 879 #SymphonyOfLoss 1  
## 880 #SymphonyOfSpeed 1  
## 881 #TalentShow 1  
## 882 #TapestryOfUnderstanding 1  
## 883 #TaylorSwift 1  
## 884 #TeamSpirit 1  
## 885 #Teamwork 1  
## 886 #Tears 1  
## 887 #TeaTime 1  
## 888 #Tech 1  
## 889 #TechConference 1  
## 890 #TechIssues 1  
## 891 #TeenCartoons 1  
## 892 #TeenChef 1  
## 893 #TeenConcert 1  
## 894 #TeenConfessions 1  
## 895 #TeenCook 1  
## 896 #TeenCreative 1  
## 897 #TeenCuriosity 1  
## 898 #TeenExcitement 1  
## 899 #TeenFashion 1  
## 900 #TeenGamer 1  
## 901 #TeenMusic 1  
## 902 #TeenPlaylist 1  
## 903 #TeenProblems 1  
## 904 #TeenReading 1  
## 905 #TeenStyle 1  
## 906 #TeenSuccess 1  
## 907 #TeenThoughts 1  
## 908 #TeenTravel 1  
## 909 #TennisRivalry 1  
## 910 #TennisSetback 1  
## 911 #TextbookJuggling 1  
## 912 #TextingFail 1  
## 913 #ThankfulMelody 1  
## 914 #ThrillerMovie 1  
## 915 #ThrillingJourney 1  
## 916 #TightropeWalk 1  
## 917 #TimeCapsule 1  
## 918 #TimelessCity 1  
## 919 #TornApart 1  
## 920 #Touched 1  
## 921 #Tournament 1  
## 922 #Traffic 1  
## 923 #TranquilGarden 1  
## 924 #TranquilSunset 1  
## 925 #TravelAdventure 1  
## 926 #TravelPlans 1  
## 927 #TreasuresOfDiscovery 1  
## 928 #Triumph 1  
## 929 #U2Anthem 1  
## 930 #UncertainChoices 1  
## 931 #Unfairness 1  
## 932 #Unforgiving 1  
## 933 #Unknown 1  
## 934 #UnparalleledHeights 1  
## 935 #UnpredictableArt 1  
## 936 #UnraveledHope 1  
## 937 #UnsolvablePuzzle 1  
## 938 #UntoldSorrows 1  
## 939 #UrbanEnergy 1  
## 940 #VacationCountdown 1  
## 941 #Vibrancy 1  
## 942 #Victory 1  
## 943 #VirtualEntertainment 1  
## 944 #Volunteers 1  
## 945 #VolunteerWork 1  
## 946 #VR 1  
## 947 #VRMeetup 1  
## 948 #WarmTea 1  
## 949 #WebSeriesJourney 1  
## 950 #Weekend 1  
## 951 #WeekendAdventure 1  
## 952 #WeekendOops 1  
## 953 #WeightliftingFailure 1  
## 954 #WhisperedLoss 1  
## 955 #WhispersInTheBreeze 1  
## 956 #WhispersOfThePast 1  
## 957 #WildHeartRide 1  
## 958 #WindsOfBrokenPromises 1  
## 959 #WinterBlues 1  
## 960 #WinterMagic 1  
## 961 #WinterSports 1  
## 962 #WinterWarmth 1  
## 963 #WistfulHeart 1  
## 964 #WistfulYearning 1  
## 965 #WonderfulMoment 1  
## 966 #Wonderment 1  
## 967 #WorkFromHome 1  
## 968 #Workout 1  
## 969 #Workshop 1  
## 970 #WorldCupTriumph 1  
## 971 #Worry 1  
## 972 #WoundedHeart 1  
## 973 #YearbookMemories 1  
## 974 #YearningForConnection 1  
## 975 #ZenWisdom 1

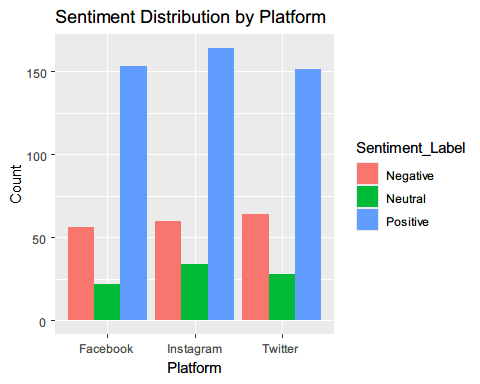
wordcloud(words = sorted\_hashtag\_counts$Hashtag,   
 freq = sorted\_hashtag\_counts$Frequency,   
 scale = c(1.5, 0.1),   
 min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35,   
 colors = brewer.pal(8, "Dark2"))



## Question 1: What are the prevailing sentiments and how do they vary across different social media platforms?

Authored by: Emma Zhang

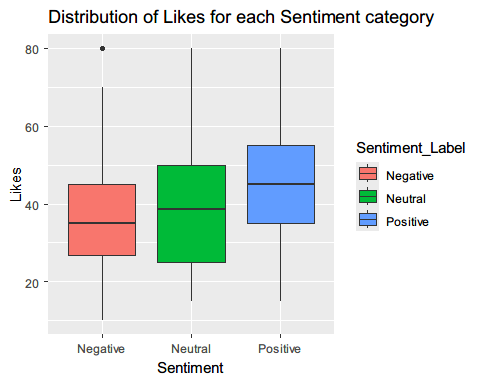
ggplot(sentiment\_platform, aes(x = trimws(Platform), y = Count, fill = Sentiment\_Label)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 labs(title = "Sentiment Distribution by Platform", x = "Platform", y = "Count")



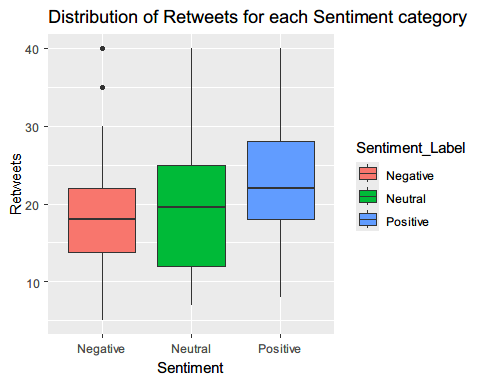
## Question 2: How do public sentiments correlate with variations in engagement metrics like likes and retweets?

Authored by: Emma Zhang

ggplot(df, aes(x = Sentiment\_Label, y = Likes, fill = Sentiment\_Label)) +  
 geom\_boxplot() +  
 labs(title = "Distribution of Likes for each Sentiment category", x = "Sentiment", y = "Likes")



ggplot(df, aes(x = Sentiment\_Label, y = Retweets, fill = Sentiment\_Label)) +  
 geom\_boxplot() +  
 labs(title = "Distribution of Retweets for each Sentiment category", x = "Sentiment", y = "Retweets")



## Question 3: Which specific keywords or phrases are most frequently associated with positive or negative sentiments?

Authored by: Xiaoxin Bian

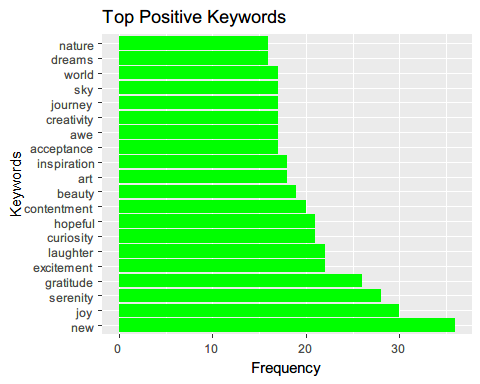
# Cleaning and splitting text  
text\_data <- df %>%  
 mutate(Text = tolower(Text)) %>%  
 unnest\_tokens(word, Text)  
  
# Cleaning and splitting hashtags  
hashtag\_data <- df %>%  
 mutate(Hashtags = str\_replace\_all(Hashtags, " #", ",")) %>%  
 separate\_rows(Hashtags, sep = ",") %>%  
 filter(Hashtags != "") %>%  
 mutate(word = tolower(Hashtags)) # Create a uniform column name for merging  
  
# Combine both datasets  
combined\_data <- bind\_rows(text\_data, hashtag\_data)

stop\_words <- get\_stopwords()  
  
cleaned\_data <- combined\_data %>%  
 anti\_join(stop\_words, by = "word")

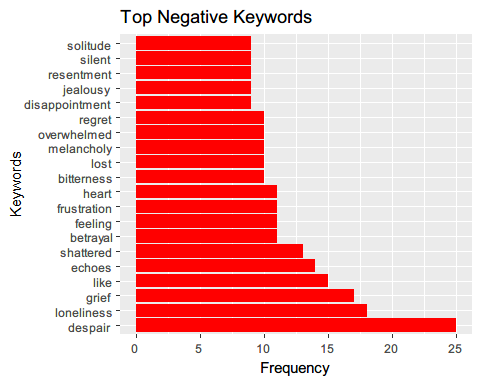
keyword\_sentiment <- cleaned\_data %>%  
 group\_by(word, Sentiment\_Label) %>%  
 summarise(Count = n(), .groups = 'drop')  
  
print(keyword\_sentiment)

## # A tibble: 3,897 × 3  
## word Sentiment\_Label Count  
## <chr> <chr> <int>  
## 1 "1" Positive 1  
## 2 "ablaze" Positive 1  
## 3 "abstract" Negative 1  
## 4 "abstract" Neutral 1  
## 5 "abstractart " Neutral 1  
## 6 "abyss" Negative 4  
## 7 "abyssofheartache " Negative 1  
## 8 "academic" Negative 1  
## 9 "academic" Positive 1  
## 10 "academicfrustration" Negative 1  
## # ℹ 3,887 more rows

# Filter for positive and negative sentiments  
positive\_words <- keyword\_sentiment %>%  
 filter(Sentiment\_Label == "Positive") %>%  
 arrange(desc(Count))  
  
negative\_words <- keyword\_sentiment %>%  
 filter(Sentiment\_Label == "Negative") %>%  
 arrange(desc(Count))  
  
# Use ggplot2 for visualization  
library(ggplot2)  
ggplot(positive\_words[1:20,], aes(x = reorder(word, -Count), y = Count)) +  
 geom\_bar(stat = "identity", fill = "green") +  
 labs(title = "Top Positive Keywords", x = "Keywords", y = "Frequency") +  
 coord\_flip()



ggplot(negative\_words[1:20,], aes(x = reorder(word, -Count), y = Count)) +  
 geom\_bar(stat = "identity", fill = "red") +  
 labs(title = "Top Negative Keywords", x = "Keywords", y = "Frequency") +  
 coord\_flip()



## Question 4: Develop machine learning models that can predict the sentiment of a text based on the keywords it contains.

Authored by: Xiaoxin Bian

# Check and clean data, handle NA values if necessary  
data <- na.omit(df)  
  
# Create a corpus and clean text data  
corpus <- VCorpus(VectorSource(data$Text))  
corpus <- tm\_map(corpus, content\_transformer(tolower))  
corpus <- tm\_map(corpus, removePunctuation)  
corpus <- tm\_map(corpus, removeNumbers)  
corpus <- tm\_map(corpus, removeWords, stopwords("english"))  
  
# Create a document-term matrix  
dtm <- DocumentTermMatrix(corpus)  
data\_matrix <- as.matrix(dtm)  
colnames(data\_matrix) <- make.names(colnames(data\_matrix))

set.seed(123)  
training\_samples <- createDataPartition(data$Sentiment\_Label, p = 0.8, list = FALSE)  
  
train\_data <- data\_matrix[training\_samples, ]  
test\_data <- data\_matrix[-training\_samples, ]  
train\_labels <- data$Sentiment\_Label[training\_samples]  
test\_labels <- data$Sentiment\_Label[-training\_samples]

# Assuming your labels are 'Negative', 'Neutral', 'Positive'  
all\_levels <- c('Negative', 'Neutral', 'Positive')  
  
# Convert labels to factors making sure all levels are included  
train\_labels <- factor(train\_labels, levels = all\_levels)  
test\_labels <- factor(test\_labels, levels = all\_levels)

# Fit a SVM model  
model\_svm <- train(train\_data, train\_labels, method = "svmRadial", trControl = trainControl(method = "cv", number = 10))

## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.  
## Warning in .local(x, ...): Variable(s) `' constant. Cannot scale data.

# Print the model summary  
print(model\_svm)

## Support Vector Machines with Radial Basis Function Kernel   
##   
## 587 samples  
## 2468 predictors  
## 3 classes: 'Negative', 'Neutral', 'Positive'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 528, 528, 527, 528, 527, 529, ...   
## Resampling results across tuning parameters:  
##   
## C Accuracy Kappa   
## 0.25 0.6389362 0.000000000  
## 0.50 0.6406029 0.006180194  
## 1.00 0.6866395 0.178506006  
##   
## Tuning parameter 'sigma' was held constant at a value of 0.06719368  
## Accuracy was used to select the optimal model using the largest value.  
## The final values used for the model were sigma = 0.06719368 and C = 1.

# Predict on test data  
predictions <- predict(model\_svm, test\_data)  
  
# Generate a confusion matrix  
conf\_matrix <-confusionMatrix(predictions, test\_labels)

## Question 5: Are there notable differences in sentiment and engagement when comparing user-generated content across different regions? How do these differences impact overall sentiment trends? [Geographical Strategy]

Authored by: Ricky Wong

In marketing trend investigations, understanding regional differences in user-generated content is crucial for tailoring strategies to specific markets. Social media users come from diverse regions, and analyzing sentiment and engagement metrics across these regions helps us optimize our marketing efforts.

By examining our dataset, which categorizes user-generated content by sentiment (positive, negative, neutral) and includes engagement metrics like likes and retweets, we can identify regional trends. This analysis focuses on North America and Europe to highlight key differences.

Analysis Strategy: 1. We grouped the data by country and sentiment to observe variations. We focused on the top 10 countries by post count, calculating the percentage of each sentiment. 2. We want to focus in North America and Europe, we grouped the top 10 countries into North America and Europe. We then again analyze sentiment distribution within these regions. 3. For each region, identify the top 10 hashtags based on the total number of retweets from posts containing these hashtags.

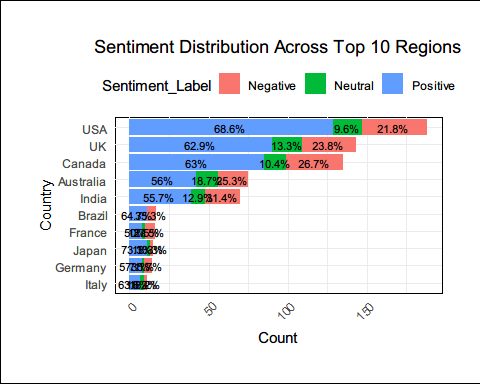
sentiment\_by\_region <- df %>%  
 group\_by(Country, Sentiment\_Label) %>%  
 summarize(count = n()) %>%  
 ungroup()

## `summarise()` has grouped output by 'Country'. You can override using the  
## `.groups` argument.

top\_countries <- sentiment\_by\_region %>%  
 group\_by(Country) %>%  
 summarize(total\_count = sum(count)) %>%  
 top\_n(10, total\_count) %>%  
 pull(Country)  
  
top\_sentiment\_by\_region <- sentiment\_by\_region %>%  
 filter(Country %in% top\_countries)  
  
top\_sentiment\_by\_region <- top\_sentiment\_by\_region %>%  
 group\_by(Country) %>%  
 mutate(percentage = count / sum(count) \* 100) %>%  
 ungroup()  
  
top\_sentiment\_by\_region <- top\_sentiment\_by\_region %>%  
 left\_join(  
 sentiment\_by\_region %>%  
 group\_by(Country) %>%  
 summarize(total\_count = sum(count)),  
 by = "Country"  
 )  
  
top\_sentiment\_by\_region

## # A tibble: 29 × 5  
## Country Sentiment\_Label count percentage total\_count  
## <fct> <chr> <int> <dbl> <int>  
## 1 Australia Negative 19 25.3 75  
## 2 Australia Neutral 14 18.7 75  
## 3 Australia Positive 42 56 75  
## 4 Brazil Negative 6 35.3 17  
## 5 Brazil Positive 11 64.7 17  
## 6 Canada Negative 36 26.7 135  
## 7 Canada Neutral 14 10.4 135  
## 8 Canada Positive 85 63.0 135  
## 9 France Negative 6 37.5 16  
## 10 France Neutral 2 12.5 16  
## # ℹ 19 more rows

p1 <- ggplot(top\_sentiment\_by\_region, aes(x = reorder(Country, total\_count), y = count, fill = Sentiment\_Label)) +  
 geom\_bar(stat = "identity", position = "stack") +  
 geom\_text(aes(label = paste0(round(percentage, 1), "%")),   
 position = position\_stack(vjust = 0.5), size = 3) +  
 ggtitle("Sentiment Distribution Across Top 10 Regions") +  
 theme\_minimal() +  
 xlab("Country") +  
 ylab("Count") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1),  
 plot.title = element\_text(hjust = 0.5),  
 legend.position = "top") +  
 theme(plot.margin = margin(5, 5, 5, 5),  
 plot.background = element\_rect(fill = "white"),  
 panel.background = element\_rect(fill = "white")) +  
 theme(plot.margin = unit(c(1, 1, 1, 1), "cm")) +  
 coord\_flip() +  
 theme(plot.background = element\_rect(fill = "white"),  
 panel.background = element\_rect(fill = "white")) +  
 theme(plot.margin = unit(c(1, 1, 1, 1), "cm"))  
  
p1

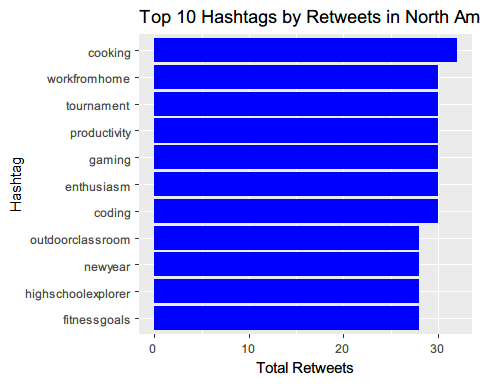


# ggsave("./charts/sentiment\_distribution\_across\_regions.png", p1, height = 15, width =25)

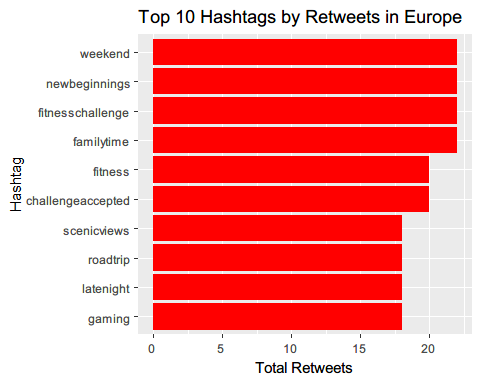
### Results 1

From the analysis, it is observed that the USA has the highest number of posts. Among these posts, approximately 70% express a positive sentiment, which is the highest compared to the other nine countries. In the top 10 countries, North America includes the USA and Canada, while European countries include the UK, France, Germany, and Italy.

north\_america <- c("USA", "Canada")  
europe <- c("UK", "Germany", "France", "Italy")  
  
df\_na <- df %>% filter(Country %in% north\_america)  
df\_eu <- df %>% filter(Country %in% europe)  
  
# Top 10 hashtags by retweets in North America  
top\_hashtags\_na <- df\_na %>%  
 filter(Sentiment == "Positive") %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(total\_retweets = sum(Retweets)) %>%  
 top\_n(10, total\_retweets)  
  
# Top 10 hashtags by retweets in Europe  
top\_hashtags\_eu <- df\_eu %>%  
 filter(Sentiment == "Positive") %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(total\_retweets = sum(Retweets)) %>%  
 top\_n(10, total\_retweets)  
  
# Plot top hashtags by retweets for North America  
p2 <- ggplot(top\_hashtags\_na, aes(x = reorder(word, total\_retweets), y = total\_retweets)) +  
 geom\_bar(stat = "identity", fill = "blue") +  
 coord\_flip() +  
 ggtitle("Top 10 Hashtags by Retweets in North America") +  
 xlab("Hashtag") +  
 ylab("Total Retweets")  
  
print(p2)



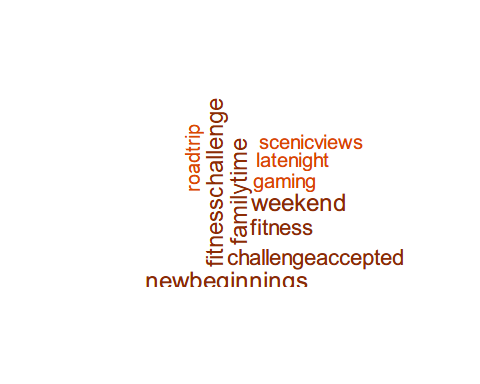
# Plot top hashtags by retweets for Europe  
p3 <- ggplot(top\_hashtags\_eu, aes(x = reorder(word, total\_retweets), y = total\_retweets)) +  
 geom\_bar(stat = "identity", fill = "red") +  
 coord\_flip() +  
 ggtitle("Top 10 Hashtags by Retweets in Europe") +  
 xlab("Hashtag") +  
 ylab("Total Retweets")  
  
print(p3)



# Create word clouds for top hashtags  
wordcloud(words = top\_hashtags\_na$word, freq = top\_hashtags\_na$total\_retweets,   
 scale = c(1.5, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35,colors = brewer.pal(8, "Blues"), main = "Top 3 Hashtags in North America")



wordcloud(words = top\_hashtags\_eu$word, freq = top\_hashtags\_eu$total\_retweets,   
 scale = c(1.5, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35,colors = brewer.pal(8, "Oranges"), main = "Top 3 Hashtags in Europe")



### Results 2

We aggregated the data according to the countries belonging to North America and Europe and further analyzed it. We identified the top 10 hashtags by retweets in both regions. For North America, the top hashtags are #cooking, #workfromhome, #productivity, #coding, and so on. For Europe, the leading hashtags are #fitnesschallenge, #familytime, #weekend, and #newbeginnings.

### Insights

Understanding these regional differences allows us to:

* tailor content to fit the preferences and sentiments of each region, improving engagement and conversion rates. For example, …
* use sentiment and hashtag trends to create targeted marketing campaigns, such as leveraging nostalgic content in regions where it performs well. For example, …

## Question 6: How do the sentiments associated with key marketing terms change over time? Which months have the highest counts of positive and negative sentiments, and what are the trending hashtags during these periods? [Temporal Trend strategy]

Authored by: Ricky Wong

df$Timestamp <- as.POSIXct(df$Timestamp, format="%Y-%m-%d %H:%M:%S")  
  
# Extract year and month for time series analysis  
df <- df %>%  
 mutate(YearMonth = format(Timestamp, "%Y-%m"))  
  
# Filter data to include only from 2020 onward  
df\_2020 <- df %>%  
 filter(format(Timestamp, "%Y") >= 2020)  
  
# Group by YearMonth and Sentiment to get the count of each sentiment per month  
sentiment\_over\_time <- df\_2020 %>%  
 group\_by(YearMonth, Sentiment\_Label) %>%  
 summarize(count = n()) %>%  
 ungroup()

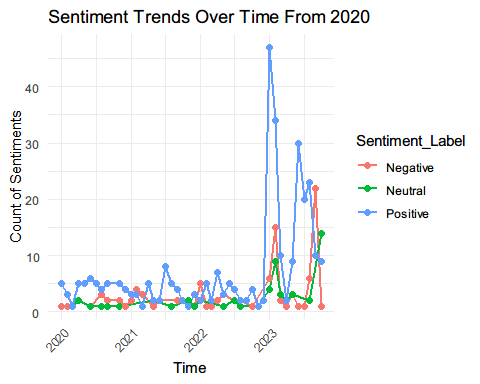
## `summarise()` has grouped output by 'YearMonth'. You can override using the  
## `.groups` argument.

sentiment\_over\_time

## # A tibble: 103 × 3  
## YearMonth Sentiment\_Label count  
## <chr> <chr> <int>  
## 1 2020-01 Negative 1  
## 2 2020-01 Positive 5  
## 3 2020-02 Negative 1  
## 4 2020-02 Positive 3  
## 5 2020-03 Negative 1  
## 6 2020-03 Neutral 1  
## 7 2020-03 Positive 1  
## 8 2020-04 Negative 2  
## 9 2020-04 Neutral 2  
## 10 2020-04 Positive 5  
## # ℹ 93 more rows

# Plot the sentiment trends over time  
ggplot(sentiment\_over\_time, aes(x = as.Date(paste0(YearMonth, "-01")), y = count, color = Sentiment\_Label)) +  
 geom\_line(size = 1) +  
 geom\_point(size = 2) +  
 ggtitle("Sentiment Trends Over Time From 2020") +  
 xlab("Time") +  
 ylab("Count of Sentiments") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



max\_positive\_month <- sentiment\_over\_time %>%  
 filter(Sentiment\_Label == "Positive") %>%  
 arrange(desc(count)) %>%  
 slice(1) %>%  
 pull(YearMonth)  
  
max\_negative\_month <- sentiment\_over\_time %>%  
 filter(Sentiment\_Label == "Negative") %>%  
 arrange(desc(count)) %>%  
 slice(1) %>%  
 pull(YearMonth)  
  
positive\_month\_data <- df %>%  
 filter(YearMonth == max\_positive\_month)  
  
negative\_month\_data <- df %>%  
 filter(YearMonth == max\_negative\_month)  
  
top\_positive\_hashtags <- positive\_month\_data %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(post\_count = n()) %>%  
 arrange(desc(post\_count))  
  
top\_negative\_hashtags <- negative\_month\_data %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(post\_count = n()) %>%  
 arrange(desc(post\_count))  
  
top\_positive\_hashtags

## # A tibble: 105 × 2  
## word post\_count  
## <chr> <int>  
## 1 cooking 3  
## 2 art 2  
## 3 fitness 2  
## 4 gaming 2  
## 5 nature 2  
## 6 reading 2  
## 7 reflection 2  
## 8 virtualreality 2  
## 9 acceptance 1  
## 10 accomplished 1  
## # ℹ 95 more rows

top\_negative\_hashtags

## # A tibble: 58 × 2  
## word post\_count  
## <chr> <int>  
## 1 highschoolstruggles 3  
## 2 empathetic 2  
## 3 highschoolemotions 2  
## 4 highschoolreality 2  
## 5 lanternofcompassion 2  
## 6 academicfrustration 1  
## 7 badday 1  
## 8 badhairday 1  
## 9 badluck 1  
## 10 badluckday 1  
## # ℹ 48 more rows

wordcloud(words = top\_positive\_hashtags$word, freq = top\_positive\_hashtags$post\_count,   
 scale = c(2, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35, colors = brewer.pal(8, "Blues"), main = paste("Top Hashtags in", max\_positive\_month, "(Positive)"))

## Warning in wordcloud(words = top\_positive\_hashtags$word, freq =  
## top\_positive\_hashtags$post\_count, : vacationcountdown could not be fit on page.  
## It will not be plotted.



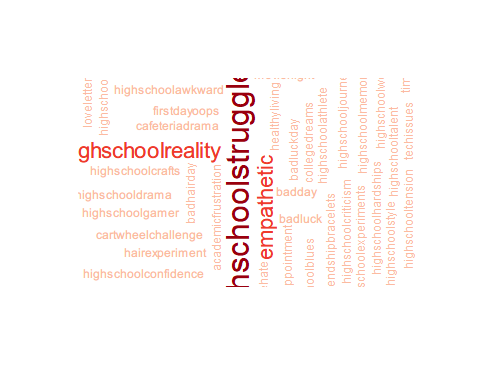
wordcloud(words = top\_negative\_hashtags$word, freq = top\_negative\_hashtags$post\_count,   
 scale = c(2, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35, colors = brewer.pal(8, "Reds"), main = paste("Top Hashtags in", max\_negative\_month, "(Negative)"))

## Warning in wordcloud(words = top\_negative\_hashtags$word, freq =  
## top\_negative\_hashtags$post\_count, : projectstruggles could not be fit on page.  
## It will not be plotted.

## Warning in wordcloud(words = top\_negative\_hashtags$word, freq =  
## top\_negative\_hashtags$post\_count, : sayinggoodbye could not be fit on page. It  
## will not be plotted.

## Warning in wordcloud(words = top\_negative\_hashtags$word, freq =  
## top\_negative\_hashtags$post\_count, : sciencegeek could not be fit on page. It  
## will not be plotted.

## Warning in wordcloud(words = top\_negative\_hashtags$word, freq =  
## top\_negative\_hashtags$post\_count, : standupcomedy could not be fit on page. It  
## will not be plotted.



### Results

From the analysis, we observed that the months with the highest counts of positive and negative sentiments vary. Specifically, January 2023 had the highest count of positive sentiment, while September 2023 had the highest count of negative sentiment.

For the month with the highest positive sentiment, the top hashtags included:

* #cooking
* #nature
* #gaming
* #fitness

For the month with the highest negative sentiment, the top hashtags included:

* #highschoolstruggles
* #highschoolreality
* #highschoolemotions

This pattern is quite reasonable and understandable. January, being the start of the year, is a time when people tend to stay positive and motivated, setting new goals. On the other hand, September marks the beginning of a new academic year. Adolescents face new challenges and uncertainties, especially those entering high school, which can lead to feelings of pessimism or negativity.

### Insight

Understanding these temporal sentiment trends enables Ignite Digital to tailor their marketing strategies more effectively. For example, promoting wellness and self-improvement content in January or addressing back-to-school anxieties in September can lead to more impactful and timely marketing efforts. By aligningour campaigns with the flow of public sentiment, we can enhance engagement and resonance with our audience.

## Shiny Dashboard

library(shiny)  
library(shinydashboard)  
library(ggplot2)  
library(dplyr)  
library(tidytext)  
  
ui <- dashboardPage(  
 dashboardHeader(title = "Sentiment Analysis Dashboard"),  
 dashboardSidebar(  
 sidebarMenu(  
 menuItem("Question 1", tabName = "question1", icon = icon("chart-column")),  
 menuItem("Question 2", tabName = "question2", icon = icon("chart-simple")),  
 menuItem("Question 3", tabName = "question3", icon = icon("chart-bar")),  
 menuItem("Question 4", tabName = "question4", icon = icon("cog")),   
 menuItem("Question 5", tabName = "question5", icon = icon("chart-bar")),  
 menuItem("Question 6", tabName = "question6", icon = icon("line-chart"))  
 )  
 ),  
 dashboardBody(  
 useShinyjs(),  
 tabItems(  
 tabItem(tabName = "question1",  
 fluidRow(  
 box(  
 title = "Question 1",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("What are the prevailing sentiments and how do they vary across different social media platforms?"),  
 p("Authored by: Emma Zhang"),  
 p("Knowing the sentiment on each platform can help tailor content and engagement strategies."),  
 p("Across all three platforms, positive sentiments are significantly higher compared to neutral and negative sentiments. This suggests that users generally express more positive emotions on the social media.")  
 ),  
 box(  
 title = "Sentiment Distribution by Platform", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("sentimentDistribution", height = 600, click = "plot\_click"),  
 width = 12  
 ),  
 )),  
 tabItem(tabName = "question2",  
 fluidRow(  
 box(  
 title = "Question 2",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("How do public sentiments correlate with variations in engagement metrics like likes and retweets?"),  
 p("Authored by: Emma Zhang"),  
 p("Engagement metrics like likes and retweets are indicators of how well content is performing. Understanding their correlation with sentiment can help optimize content to achieve higher engagement, thereby increasing the visibility and reach of posts."),  
 p("Positive posts receive the highest median engagement for both likes and retweets. Users are more likely to share and like positive content."),  
 p("Neutral posts receive consistent engagement across likes and retweets, indicating that users engage with such content without significant spikes or drops."),  
 p("Negative posts receive the lowest median engagement for both likes and retweets. But there are a few outliers in both metrics. The higher engagement might be due to the controversial or attention-grabbing nature of negative content."),  
 ),  
   
 box(  
 title = "Distribution of Likes for each Sentiment category", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("likesDistribution", height = 600, click = "plot\_click"),  
 width = 12  
 ),  
 box(  
 title = "Distribution of Retweets for each Sentiment category", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("retweentsDistribution", height = 600, click = "plot\_click"),  
 width = 12  
 ),  
 )),  
 tabItem(tabName = "question3",   
 fluidRow(  
 box(  
 title = "Question 3",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("Which specific keywords or phrases are most frequently associated with positive or negative sentiments?"),  
 p("Authored by: Xiaoxin Bian"),  
 p("The test and hashtag was processed to extract individual words and their associated sentiment labels."),  
 p("Understanding the specific keywords associated with positive or negative sentiments can help in tailoring content to evoke desired emotions and improve engagement.")  
 ),  
 box(  
 title = "Top Positive Keywords", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("positiveKeywordsPlot", height = 600), # Output for positive keywords plot  
 width = 12  
 ),  
 box(  
 title = "Top Negative Keywords", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("negativeKeywordsPlot", height = 600), # Output for negative keywords plot  
 width = 12  
 ),  
 )),  
 tabItem(tabName = "question4",   
 fluidRow(  
 box(  
 title = "Question 4",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("Develop machine learning models that can predict the sentiment of a text based on the keywords it contains."),  
 p("Authored by: Xiaoxin Bian"),  
 p("The objective is to build a machine learning model to predict the sentiment of a text based on the keywords it contains. The Support Vector Machine (SVM) model was trained using a document-term matrix representation of the text data."),  
 p("The model achieved an accuracy of 71.03% with the following confusion matrix:")  
 ),  
 box(  
 title = "Confusion Matrix", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 tableOutput("confMatrix"), # Output for confusion matrix  
 width = 12  
 ),  
 box(  
 title = "Model Summary", status = "primary",   
 solidHeader = TRUE, collapsible = TRUE,  
 verbatimTextOutput("modelSummary"), # Output for model summary  
 width = 12  
 )  
 )),  
 tabItem(tabName = "question5",  
 fluidRow(  
 box(  
 title = "Question 5",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("Are there notable differences in sentiment and engagement when comparing user-generated content across different regions? How do these differences impact overall sentiment trends? [Geographical Strategy]"),  
 p("Authored by: Ricky Wong"),  
 p("In marketing trend investigations, understanding regional differences in user-generated content is crucial for tailoring strategies to specific markets. Social media users come from diverse regions, and analyzing sentiment and engagement metrics across these regions helps us optimize our marketing efforts."),  
 p("By examining our dataset, which categorizes user-generated content by sentiment (positive, negative, neutral) and includes engagement metrics like likes and retweets, we can identify regional trends. This analysis focuses on North America and Europe to highlight key differences."),  
 p("Analysis Strategy:"),  
 p("1. We grouped the data by country and sentiment to observe variations. We focused on the top 10 countries by post count, calculating the percentage of each sentiment."),  
 p("2. We want to focus in North America and Europe, we grouped the top 10 countries into North America and Europe. We then again analyze sentiment distribution within these regions."),  
 p("3. For each region, identify the top 10 hashtags based on the total number of retweets from posts containing these hashtags.")  
 ),  
 box(  
 title = "Sentiment Distribution Across Top 10 Regions", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("sentimentPlot", height = 600, click = "plot\_click"),  
 width = 12  
 ),  
 box(  
 title = "Top Hashtags by Retweets", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 selectInput("region", "Select Region:", choices = c("North America", "Europe")),  
 plotOutput("hashtagPlot", height = 600),  
 dataTableOutput("hashtagTable"),  
 width = 6  
 ),  
 box(  
 title = "Word Cloud of Top Hashtags", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("wordcloudPlot", height = 600),  
 width = 6  
 )  
 )  
 ),  
   
 # Question 6 tab content  
 tabItem(tabName = "question6",  
 fluidRow(  
 box(  
 title = "Question 6",  
 status = "primary",  
 solidHeader = TRUE,  
 collapsible = TRUE,  
 width = 12,  
 h4("How do the sentiments associated with key marketing terms change over time? Which months have the highest counts of positive and negative sentiments, and what are the trending hashtags during these periods? [Temporal Trend strategy]"),  
 p("Authored by: Ricky Wong"),  
 p("Understanding how sentiments change over time is crucial for marketers to adapt their strategies to current trends. By examining the dataset over a specific time range, we can identify patterns and peak periods of positive and negative sentiments."),  
 p("Analysis Strategy:"),  
 p("1. We extracted the year and month from the timestamp of each post to analyze sentiment trends over time."),  
 p("2. We filtered the data to include only posts from 2020 onward."),  
 p("3. We grouped the data by year-month and sentiment to calculate the count of each sentiment per month."),  
 p("4. For the months with the highest counts of positive and negative sentiments, we identified the top hashtags based on post count.")  
 ),  
 box(  
 title = "Sentiment Trends Over Time", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 dateRangeInput("dateRange", "Select Date Range:",  
 start = min(df$Timestamp), end = max(df$Timestamp),  
 min = min(df$Timestamp), max = max(df$Timestamp)),  
 plotOutput("trendPlot", height = 600),  
 width = 12  
 ),  
 box(  
 title = "Top Hashtags in Months with Highest Positive Sentiments", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("positiveWordcloudPlot", height = 500),  
 width = 6  
 ),  
 box(  
 title = "Top Hashtags in Months with Highest Negative Sentiments", status = "primary", solidHeader = TRUE, collapsible = TRUE,  
 plotOutput("negativeWordcloudPlot", height = 500),  
 width = 6  
 )  
 )  
 )  
 )  
 )  
)  
  
# Define server logic  
server <- function(input, output) {  
   
 #Render plot for question 1  
 output$sentimentDistribution <- renderPlot({  
 ggplot(sentiment\_platform, aes(x = trimws(Platform), y = Count, fill = Sentiment\_Label)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 labs(title = "Sentiment Distribution by Platform", x = "Platform", y = "Count")  
 })  
   
 #Render plot for question 2 - Likes distribution  
 output$likesDistribution <- renderPlot({  
 ggplot(df, aes(x = Sentiment\_Label, y = Likes, fill = Sentiment\_Label)) +  
 geom\_boxplot() +  
 labs(title = "Distribution of Likes for each Sentiment category", x = "Sentiment", y = "Likes")  
 })  
   
 output$retweentsDistribution <- renderPlot({  
 ggplot(df, aes(x = Sentiment\_Label, y = Retweets, fill = Sentiment\_Label)) +  
 geom\_boxplot() +  
 labs(title = "Distribution of Retweets for each Sentiment category", x = "Sentiment",   
 y = "Retweets")  
 })  
   
 # Render positive keywords plot for Question 3  
 output$positiveKeywordsPlot <- renderPlot({  
 ggplot(positive\_words[1:20,], aes(x = reorder(word, -Count), y = Count)) +  
 geom\_bar(stat = "identity", fill = "green") +  
 labs(title = "Top Positive Keywords", x = "Keywords", y = "Frequency") +  
 coord\_flip()+  
 theme\_minimal()  
 })  
   
 output$negativeKeywordsPlot <- renderPlot({  
 ggplot(negative\_words[1:20,], aes(x = reorder(word, -Count), y = Count)) +  
 geom\_bar(stat = "identity", fill = "red") +  
 labs(title = "Top Negative Keywords", x = "Keywords", y = "Frequency") +  
 coord\_flip()+  
 theme\_minimal()  
 })  
   
 output$confMatrix <- renderTable({  
 conf\_matrix$table  
 })  
   
 output$modelSummary <- renderPrint({  
 model\_svm  
 })  
   
   
 # Render sentiment distribution plot for Question 5  
 output$sentimentPlot <- renderPlot({  
 ggplot(top\_sentiment\_by\_region, aes(x = reorder(Country, total\_count), y = count, fill = Sentiment\_Label)) +  
 geom\_bar(stat = "identity", position = "stack") +  
 geom\_text(aes(label = paste0(round(percentage, 1), "%")),   
 position = position\_stack(vjust = 0.5), size = 3) +  
 ggtitle("Sentiment Distribution Across Top 10 Regions") +  
 theme\_minimal() +  
 xlab("Country") +  
 ylab("Count") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1),  
 plot.title = element\_text(hjust = 0.5),  
 legend.position = "top",  
 plot.margin = unit(c(1,1,1,1), "cm")) +  
 coord\_flip()  
 })  
   
 # Reactive expressions for region-specific data  
 region\_data <- reactive({  
 if (input$region == "North America") {  
 df %>% filter(Country %in% north\_america)  
 } else {  
 df %>% filter(Country %in% europe)  
 }  
 })  
   
 # Render hashtag plot based on selected region for Question 5  
 output$hashtagPlot <- renderPlot({  
 data <- region\_data()  
   
 top\_hashtags <- data %>%  
 filter(Sentiment\_Label == "Positive") %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(total\_retweets = sum(Retweets)) %>%  
 top\_n(10, total\_retweets)  
   
 ggplot(top\_hashtags, aes(x = reorder(word, total\_retweets), y = total\_retweets)) +  
 geom\_bar(stat = "identity", fill = ifelse(input$region == "North America", "blue", "red")) +  
 coord\_flip() +  
 ggtitle(paste("Top 10 Hashtags by Retweets in", input$region)) +  
 xlab("Hashtag") +  
 ylab("Total Retweets")  
 })  
   
 # Render hashtag data table based on selected region for Question 5  
 output$hashtagTable <- renderDataTable({  
 data <- region\_data()  
   
 top\_hashtags <- data %>%  
 filter(Sentiment\_Label == "Positive") %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(total\_retweets = sum(Retweets))  
   
 datatable(top\_hashtags, options = list(pageLength = 10))  
 })  
   
 # Render word cloud based on selected region for Question 5  
 output$wordcloudPlot <- renderPlot({  
 data <- region\_data()  
   
 top\_hashtags <- data %>%  
 filter(Sentiment\_Label == "Positive") %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(total\_retweets = sum(Retweets))  
   
 if (nrow(top\_hashtags) > 0) {  
 wordcloud(words = top\_hashtags$word, freq = top\_hashtags$total\_retweets,   
 scale = c(2, 0.1), min.freq = 1, max.words = 200, random.order = FALSE,   
 rot.per = 0.35, colors = ifelse(input$region == "North America", brewer.pal(7,"BrBG"), brewer.pal(7, "Dark2")))  
 } else {  
 plot.new()  
 text(0.5, 0.5, "No data available for the selected region", cex = 1.5)  
 }  
   
 })  
   
 # Reactive expression to filter data based on selected date range  
 filtered\_data <- reactive({  
 df %>% filter(Timestamp >= input$dateRange[1] & Timestamp <= input$dateRange[2])  
 })  
   
 # Render sentiment trends plot for Question 6  
 output$trendPlot <- renderPlot({  
 data <- filtered\_data() %>%  
 mutate(YearMonth = format(Timestamp, "%Y-%m")) %>%  
 group\_by(YearMonth, Sentiment\_Label) %>%  
 summarize(count = n()) %>%  
 ungroup()  
   
 ggplot(data, aes(x = as.Date(paste0(YearMonth, "-01")), y = count, color = Sentiment\_Label)) +  
 geom\_line(size = 1) +  
 geom\_point(size = 2) +  
 ggtitle("Sentiment Trends Over Time") +  
 xlab("Time") +  
 ylab("Count of Sentiments") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))  
 })  
   
 # Render positive word cloud for Question 6  
 output$positiveWordcloudPlot <- renderPlot({  
 data <- filtered\_data() %>%  
 filter(YearMonth == max\_positive\_month) %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(post\_count = n())  
   
 if (nrow(data) > 0) {  
 wordcloud(words = data$word, freq = data$post\_count,   
 scale = c(2.2, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35, colors = brewer.pal(8, "Blues"), main = paste("Top Hashtags in Month with Highest Positive Sentiment"))  
 } else {  
 plot.new()  
 text(0.5, 0.5, "No data available for the selected date range", cex = 1.5)  
 }  
 })  
   
 # Render negative word cloud for Question 6  
 output$negativeWordcloudPlot <- renderPlot({  
 data <- filtered\_data() %>%  
 filter(YearMonth == max\_negative\_month) %>%  
 unnest\_tokens(word, Hashtags) %>%  
 group\_by(word) %>%  
 summarize(post\_count = n())  
   
 if (nrow(data) > 0) {  
 wordcloud(words = data$word, freq = data$post\_count,   
 scale = c(2.2, 0.1), min.freq = 1,   
 max.words = 200,   
 random.order = FALSE,   
 rot.per = 0.35, colors = brewer.pal(8, "Reds"), main = paste("Top Hashtags in Month with Highest Negative Sentiment"))  
 } else {  
 plot.new()  
 text(0.5, 0.5, "No data available for the selected date range", cex = 1.5)  
 }  
 })  
}  
  
# Run the application   
shinyApp(ui = ui, server = server)

## PhantomJS not found. You can install it with webshot::install\_phantomjs(). If it is installed, please make sure the phantomjs executable can be found via the PATH variable.