You have been provided with data provided by a content intelligence firm used by news organizations and digital publishers. The data is sourced from multiple websites and indicates whether or not an individual user finished reading a given paragraph of text. The dataset contains the following variables

1. platform – Desktop, Mobile, Tablet or Unknown
2. website_id – identificator of the website from which the content comes (1 to 9)
3. weekday – weekday when the content was read by the individual (1 to 7)
4. time – time of day when the content was read by the individual (early morning, morning, afternoon, evening, late evening)
5. post25_include_# (from 1 to 25) – a 25-topic model has been estimated on the full text of a given article. The top words associated with each topic are included at the bottom of the page. Note that these variables (post25_include) will add up to 1. When you are incorporating them into your analysis, be sure to omit one variable to avoid collinearity (in other words choose the topic used as a baseline)
6. posemo, anger, fear, sadness – The text of each paragraph has been analyzed for the following emotions: Positive Emotions, Anger, Fear, Sadness. (Normalized to zero mean unit variance)
7. stopgo – variable of interest. stopgo=1 if the user continued reading the article past the current paragraph, and stopgo=0 if the reader did not continue past the current paragraph

Topic	Top 10 Words
1	china, year, oil, country, product, chinese, energy, industry, global, price
2	study, health, brain, people, patient, drug, medic, doctor, test, body
3	company, store, sale, product, price, custom, year, market, buy, retail
4	human, war, anime, people, year, live, century, history, robot, dog
5	movie, film, charact, story, episode, time, comic, star, series, scene
6	food, eat, drink, water, cook, coffee, wine, restaurant, taste, beer
7	state, law, govern, office, case, police, report, president, court, public
8	camera, screen, design, battery, iphone, phone, display, device, video, power
9	game, week, season, yard, defense, team, play, point, pass, offense
10	work, job, school, student, college, year, university, worker, employee, program
11	build, city, air, flight, project, design, airline, space, area, plane
12	time, hand, thing, work, day, sleep, clean, water, minute, body
13	year, city, car, day, house, family, time, live, room, travel
14	money, pay, tax, cost, year, plan, insure, card, credit, save
15	book, read, work, design, art, write, image, artist, model, publish
16	discussion, material, warn, graphic, reply, unapprove, map, image, county
17	company, busy, invest, fund, firm, manage, investor, billion, year, market
18	game, team, player, play, year, state, sport, football, coach, fan
19	percent, market, year, bank, rate, price, stock, growth, expect, quarter
20	app, google, window, file, set, phone, compute, android, feature, work
21	people, time, thing, good, work, feel, person, lot, talk, help
22	video, people, music, facebook, twitter, photo, song, post, news
23	data, service, site, network, secure, number, internet, online, user, call
24	women, men, sex, girl, woman, sexual, female, man, parent, children
25	earth, space, universe, light, time, planet, image, science, scientist, system

Questions

Using the data provided, estimate an appropriate regression model that will allow you to address the following questions. Be sure to include in your answers how you arrived at your conclusions.

Hint: you can answer all questions by using only one regression model

1. (10) On which types of devices are consumers more prone to complete a paragraph, and consequently have increased reading depth?
2. (10) At what time/weekday users are more likely to finish the paragraph?
3. (10) Based on the topic model, what topics are associated with increased reading depth?
4. (10) Controlling for the topic of the article, is the presence of emotional language associated with increased or decreased reading depth?
5. (10) Describe how would you use the data like this to build a recommender system. Assume you have the identification of the user in the data. In other words, you know for each user the sequence of articles they were reading during the browsing session as well as the outcome (finished or not). You should be short but clear.