Visu

Segment type:

* This section indicates whether or not an individual was part of a larger segmented group which was surveyed exclusively for that specific question.
* We have:
  + Mobile: In the "Social Influence on Shopping" dataset from data.world, the "Mobile" segment type is used to categorize social media influencers based on the type of content they create and share on mobile-based platforms, such as Instagram or TikTok. This segment type includes influencers who primarily create and share content on mobile devices, as opposed to desktop or laptop computers.
  + Web : Overall, the "Web" segment type in the "Social Influence on Shopping" dataset is a way of categorizing social media influencers based on the type of content they create and share on their own websites or blogs, and is used to analyze how these influencers impact shopping behavior among their followers based on their web-based content.
  + Gender: The purpose of the "gender" segment type is to analyze how social media influencers of different genders impact shopping behavior among their followers. By categorizing influencers into male and female segments, researchers can identify patterns and trends in consumer behavior that are influenced by these influencers based on their gender identity.
  + University : In summary, the "university" segment type in the "Social Influence on Shopping" dataset refers to a specific group of social media influencers who have a connection to universities or colleges and is used as a standardized way of categorizing influencers based on common characteristics.
  + Custom : Overall, the "custom" segment type in the "Social Influence on Shopping" dataset is a way of categorizing social media influencers based on custom criteria or attributes and is used to create tailored segments of influencers that are relevant to specific research or marketing objectives.

Segment:

* Here you will find a description of the segment population who were surveyed for each question listed in “Question” above. We have 289 diff values

Answers:

* + Facebook
  + Instagram
  + Snapchat
  + Twitter
  + None

Is there one social media that has more influence on University / overall?

What is the correlation between social media influence and shopping behavior for customers who primarily use Instagram?

Task A.

Q.1

Analyze the effects of social media influence on shopping. Are there any detectable trends?” This question was given from the coursework. We are going to use the given database of Whatsgoodly on the social influence on shopping. We are going to analyze the overall result for each social media and the none part.

Analyze the effects of social media influence on shopping . Does the actual situation and background (family, school, etc…) of the user affect media influence and shopping behavior? Social media can influence on shopping but other parameters can also influence it. So, we are going to analyze other parameters which will be the background and the actual situation to see if it has an influence on the shopping. We are going to use the same dataset than the question 1 but we are going to focus on more specified groups.

Is there one social media that has more influence on University? We are assuming that the social media affect more the teenagers and young adults. We want to verify this statement by analyzing the influence of social media on shopping only on university. To do so, we are going to use the same dataset as question 1 and 2. We have also a new dataset “Which Social Media Millenials Care About Most” which is also from Whatsgoodly, about their “preference” of social media. We can use it to compare if the social media that is preferred in university is the same that the one that influence their shopping.

Q.2

The dataset “Social influence on Shopping” is provided by Whatsgoodly, a millennial social polling company. It allow us to access data that they get from their polls.

In the data provided, we have 5 columns which are questions, segment type, segment description, answer, count. The question is: “What social platform has influenced your online shopping most? ”. To answer this question, the data has a segment type which show if whether or not an individual was part of a larger segmented group which was surveyed exclusively for that specific question. The groups are Mobile, Web, Gender, University and Custom. It comes with a segment description which is a description of the segment population who were surveyed for each question listed in “Question” above. There are 289 unique values. We also have the answers to the questions which are the social media Snapchat, Twitter, Facebook, Instagram or none. And then we have the count and the percentage for each group surveyed.

The data types for the segment type, the segment description and answers are qualitative nominal data. The count is a quantitative discrete data and percentage are quantitative continuous data.

The strength of this data is that it easy to understand, to analyze to find a pattern and does not have null of missing data. The weakness of this data is that we don’t what is the date when the data was retrieved which is an important parameter. It is also has too many segments type so the data does not have much more and deeper information about some groups. I could be interesting to know the location of where the user did the survey from.

Q.3

The two datasets have the same origin, Whatsgoodly. They have also similar columns such as segment type, segment description, answer, count and percentage. So, we can say that it is related in some way. However, there are some problems between these two datasets that might affects the correlation. For example the dataset on influence of social media on shopping has 5 possibles answer which are Facebook, Instagram, Snapchat, Twitter and None. The dataset on the social media that millennials prefer the most has for answers: Facebook, Instagram, Linkedin and Snapchat. The missing possible answer in the second dataset which can lead to incomplete correlation. Another problem is that the second datasets in about millennials and the first dataset is about the anybody in the population. So, we are going to have incomplete correlation because we are missing some age group. We can answers all the questions with the first datasets but use of second datasets can help to have more understanding and deeper answers.

Task 2.

Q1.

Each visualization should be accompanied by a maximum of 300 words describing the design rationale, which question(s) your design would help answer and if/how your design may improve upon existing examples. By design rationale we mean: the process and principles followed in choosing the specific visualization. You should provide a rigorous rationale for your design decisions, e.g. visual encodings used and why they are appropriate for the data. These decisions include the choice of visualization type, size, colour, scale, mark and channels and other visual elements, as well as the use of sorting or other data transformations. Consider how these decisions facilitate analysis and/or communication.

For the first research question, we want to create a representation of the overall count for each social media to compare them. So, we decided to choose a pie char and each slice is the total count for one social media. So I came with this design in Figure … . The pie chart will have 5 slices representing the 4 socials media and the none answer. Inside each slice, we will have the percentageOn the right of the pie chart, we will put the legend for each slice. The color used will be the reference of the logo of each social media. For example Facebook will be a dark blue, Instagram in pink, snapchat in yellow and twitter in light blue. For the none part, I just used a color that was no already used. The user will have the possibility of hover the mouse over a slice and the slice will display all the information in terms of count. The scale will be the same for each of the slice as we sum the count for each answer. The idea with this visualisation is to permit to see easily a social media that is standing out and permit to find a trend. This will facilitate analysis and communication due to it easiness to understand and read through and the display of the information of the slice wanted.

The scale used in this visualisation would follow a similar scale to the one provided by the Nextstrain dataset. Where the time would be separated into days, the clades would define the different variants and the frequency could be given in percentages. The idea to create this particular design instead of following other more common visualisations was to clearly see the independent variants and their development over time. Compared to 2D line graphs the visualisation would provide further insight into