We at The Data Monk hold the vision to make sure everyone in the IT industry has an equal stand to work in an open domain such as analytics. Analytics is one domain where there is no formal under-graduation degree and which is achievable to anyone and everyone in the World.

We are a team of 30+ mentors who have worked in various product-based companies in India and abroad, and we have come up with this idea to provide study materials directed to help you crack any analytics interview.

Every one of us has been interviewing for at least the last 6 to 8 years for different positions like Data Scientist, Data Analysts, Business Analysts, Product Analysts, Data Engineers, and other senior roles. We understand the gap between having good knowledge and converting an interview to a top product-based company.

Rest assured that if you follow our different mediums like our blog cum questions-answer portal www.TheDataMonk.com , our youtube channel - The Data Monk, and our e-books, then you will have a very strong candidature in whichever interview you participate in.

There are many blogs that provide free study materials or questions on different analytical tools and technologies, but we concentrate mostly on the questions which are asked in an interview. We have a set of 100+ books which are available both on Amazon and on The Data Monk e-shop page

We would recommend you to explore our website, youtube channel, and e-books to understand the type of questions covered in our articles. We went for the question-answer approach both on our website as well as our e-books just because we feel that the best way to go from beginner to advance level is by practicing a lot of questions on the topic.

We have launched a series of 50 e-books on our website on all the popular as well as niche topics. Our range of material ranges from SQL, Python, and Machine Learning algorithms to ANN, CNN, PCA, etc.

We are constantly working on our product and will keep on updating it. It is very necessary to go through all the questions present in this book.

Give a rating to the book on Amazon, do provide your feedback and if you want to help us grow then please subscribe to our Youtube channel.

# Guesstimates and Case studies

**Introduction:**
Before actually diving into the case studies and getting our hands dirty. Let us know why do interviewers ask these case studies and what kind of skills are they expecting from you by asking these questions. Knowing this is very important because that is when you can make your thinking aligned to the expectations.
These questions are asked not to make you feel less of yourself or trying to degrade your motive.
These are asked to:

- To examine your curiosity in finding solutions.
- What is your way of thinking for a particular solution. How are approaching the problem.
- To discover your personality or morale sometimes when your asked about any company's case study.
- to test your ability to think logically and organize your answer.

Some of the standard procedures for approaching a case study could be the following:

A. **Listen to the question and show your enthusiasm with a positive attitude**
B. **Take notes** (This might seem like an old school method but this step couldn't be stressed more).
C. **Summarize** your question by knowing the targets or objectives**.**
D. **Ask questions to the recruiters** (this might seem unnecessary but believe me some times interviewer might intentionally miss data expecting that you would ask him for it).
E. **Hold you thought for couple of seconds to process before letting it out.**
F. **Time management (**this only comes through practise take up various case studies and ask questions yourself and answer it**)**
G. **Deal with the actual numbers for getting a numerical solution.**
H. **Finally summarize** your solution with a pinch of story-telling if possible.

In this article, we shall know about Guesstimates and case studies and what are the different types of guesstimates and case studies you might encounter in interviews.

Firstly, we will look into Guesstimates.
This is how the guesstimate module is going to look like

| Segment | Question no |
|---|---|
| **Introduction to guesstimates** | 1,2 |
| **population guesstimate** | 3 |
| **household guesstimate** | 4 |
| **Structural approach guesstimate** | 5 |

Then we shall move on to Case studies. There are lots of types in case studies that you might encounter in the industry or in interviews.

We in total have 5 case studies to cover all 4 types of case studies:

1) Case study 1: cricket data set
2) Case study 2: Singapore population dataset
3) Case study 3: Spotify data set
4) Case study 4: Analysing decline in sales of Nike.
5) Case study 5: KFC its reason to become popular and what obstacles it faced in India and how did it resolve them?

| Segment | Question no |
|---|---|
| What are case studies? | 6 |
| Exploratory case study | 7,8,9 |
| Illustrative case study | 10 |
| Cumulative case study | 12 |
| Critical instance case study | 11,13,14 |

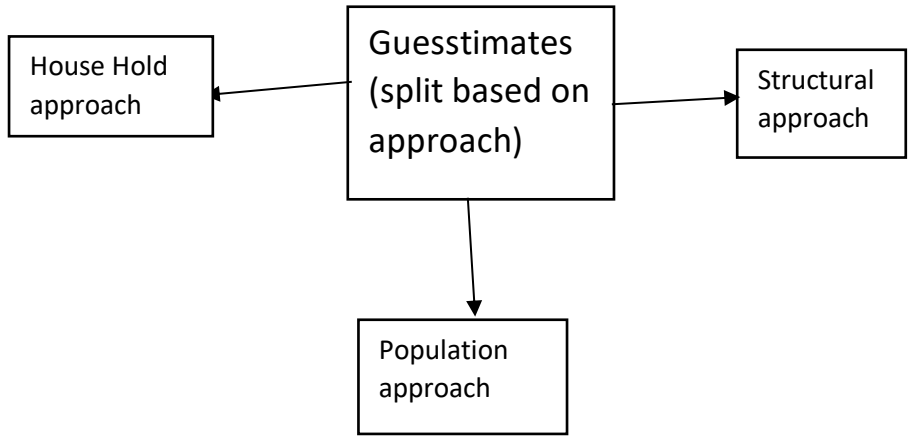
# Guesstimates

1) **What are Guesstimates?**
   Guesstimate comes from merging the two words guess and estimate.
   We define guesstimate as guessing particular numbers for a given case and form an estimation for the final random analyst problem. There are no

absolute answers for estimates, all they want to check is your approach towards getting answers.
For example: the number of chocolates eaten each day in India, how many likes are generated each day on Instagram, etc.

**2) What are different guesstimates based on approaching a solution?**

They can be broadly classified into 3 types:




We have got examples covered for each type

1) **Household approach**: The thing for which you are going to Guesstimate the number is present one for each household.
**For example**: cars, refrigerators, washing machines, dishwashers etc.

   We have an example for this method: check out the number **of vacuum cleaners in India.**

2) **Population Approach:** this is utilized or possessed by an individual person rather than a household or society.
Example: spectacles, number of people who wear black tie.
For this we have an example: find out the number **of Salman khan fans in India.**

3) **Structural approach:** this solely depends on some main factors and we find a solution by giving estimations to these factors.
The example we covered for this is the number **of sweets consumed by an average school going kid.**

After going through these 3 examples, you will be getting an idea on how to proceed with Guesstimates. Whenever you are given a guesstimate, find out to which kind this problem belongs to and then follow the procedure that is required for that kind of problem.

3) **With an example, show how the Population Guesstimate works.**

**Find out the number of Salman khan fans in India.**

Guesstimates are majorly asked to solve to know your diversified thinking. How many cases were you able to think of a given problem.
With this example let's see how we can think of solving a guesstimate.

There is a saying that if you want to know the stardom of a star, consider the collections of his flop movies not a hit film.
Going by this I would consider his flop movie to analyse my Guesstimate.

**Approach**

**Step 1: define some terms yourselves to analyse the question.**

First, Let's ask ourselves who are fans

**Fans:** I assume fans are the ones who admire the actor for his previous works or off-screen personality and would love to see him on big screen at any cost.

- Be definite in creating these own definitions; these will drive your solution.

What do I mean by **Flop**:

- The movie did not manage to get returns as much as it's budget or cost of making.
- In addition to it the movie also got bad reviews since its day of release in theatres.

Let me consider a flop movie named “XXX” acted by Salman khan.

Using this XXX movie we will find out the number of fans for him. Since we said we considered his flop movie, we assume that XXX got bad reviews all over India.

**Step 2: consider an example for the study (here I took xxx movie)**

Now, I will take the total collections of the film XXX, considering it to be 200 crores. I assume the average ticket price is 200 rupees so dividing 200 crores with 200 rupees should give me the total tickets sold. Which is 1 Crore.

**Step3: Split the audience (it need not be people, it is how you categorize things that you are asked to find, basically how do you group them)**

We now consider there are some extremely die-hard fans who dare to watch his flop movie for a second time too. Let us say there are 1 lakh people this way so the total number of people who saw the movie are 99 lakhs.

Now there are certain types of people who go to see a movie for various reasons. Let’s find out about them.

- **Category 1**: Die hand fans, who know that it got bad reviews still love to see their hero on the big screen.
- **Category2**: these people generally have a habit of going to movies, be it any hero. Common audience who are not aware of all those internet reviews and go to a nearest theatre.
- **Category 3**: People who had nothing important to do at that particular time and would go to the movies for timepass.
- **Category 4**: Since most of Salman bhai films release during the festival season. We consider some portion for people who come with their families.

Most of your skill in solving guesstimates lies here. The more meaning full splits you bring into the picture the better your guesstimate is.

The other splitting could be with dates.

Movies mostly release on Friday so that fans watch it on the first day and the following weekends can be utilized by the general audience.

**Friday:**

Of the whole audience (99 lakhs) Let's assume 40 percent of them watched a movie on the first day (the below table is about those 40 percent which is 39.6 lakhs.

| | |
|---|---|
| Category 1(fans) | 60 percent |
| Category 2(General audience) | 30 percent |
| Category 3(time pass audience) | 7 percent |
| Category 4(family audience) | 3 percent |

We only need to take category 1 so it is around (0.6*39.6) 23.76 lakhs.

**Saturday and Sunday (weekends):** these dates are filled mostly with other categories too. Assuming 50 percent of the total 99 lakhs watch on these days (49.5 lakhs).

| | |
|---|---|
| Category 1 (fans) | 15 percent |
| Category 2 (general audience) | 40 percent |
| Category 3 (time pass audience) | 20 percent |
| Category 4 (family audience) | 25 percent |

Fan count on weekends: 0.15*49.5 =7.42 lakhs

**Rest of the days till the end of the movie run in theatres**

We are left with 10 percent (9.9 lakhs), these days see good number of family audience

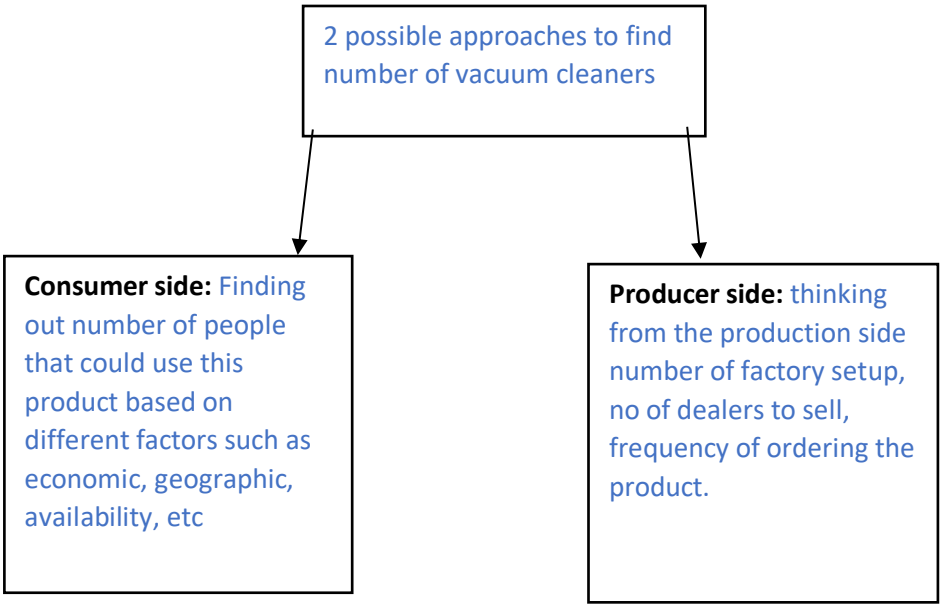
| | |
|---|---|
| Category 1 (fans) | 5 percent |
| Category 2 (general audience) | 20 percent |
| Category 3 (time pass audience) | 30 percent |
| Category 4 (family audience) | 45 percent |

Fan count: 0.05* 9.9 lakhs =0.495 lakhs

**Total fan count**: 23.76+7.42+0.495 = 31 lakh (3 million appx) fans for Salman khan

*These numbers need not reflect actual reality.

4) **Find out number of vacuum cleaners in India?(household approach)**
There can be two approaches for this problem or perse any business problem you face




You are free to take any side but be sure that you have a proper reason to support your side.

In this example I will choose Consumer side:

**Reason:** considering from the factory side these numbers could be delicate even if one factory establishes these numbers to affect all other production warehouses. We also need to take lot of assumptions if we consider production side for this example

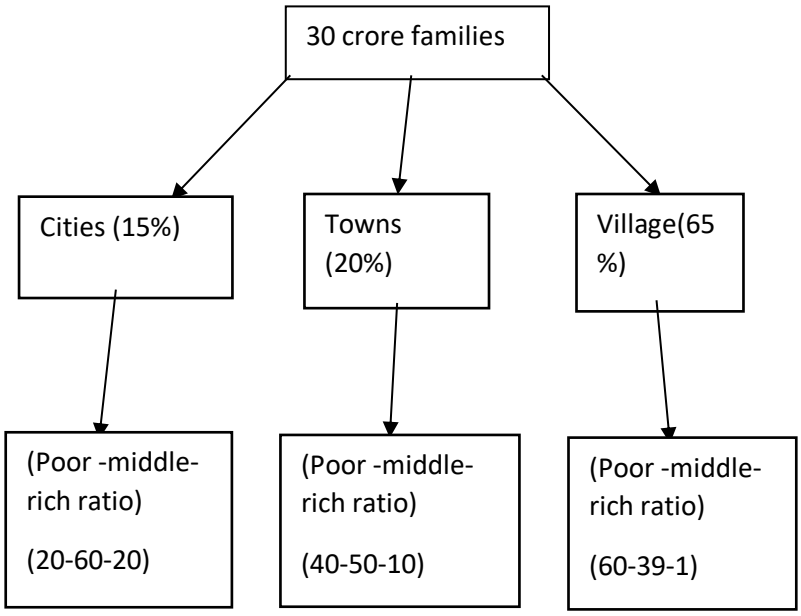
**Tip:** choose the side which needs less number of assumptions.

Choose one factor which impacts the number of products over the populations.

Total population of India is 120 crores. I will assume one product per one household so the total number of possible households is 120/4 = 30 crore (assuming 4 people per household).

These 30 crores families are divided into 3 based on the city/town/village they live in

And then they are further classified into their financial conditions such as poor, middle class, rich.




Now, we need to think of how the vacuum cleaners are distributed among these economical backgrounds.

| | **Cities** | **Towns** | **Villages** |
|---|---|---|---|
| **Poor** | 0 | 0 | 0 |
| **Middle-class** | 30 | 20 | 0 |
| **Rich** | 100 | 100 | 100 |

From the two tables given above we can find out the number of Vacuum cleaners in India.

Total vacuum cleaners would be = ∑(each sector from table 2)* (its respective population).

| | **Cities** | **Towns** | **Villages** |
|---|---|---|---|
| **Total houses** | 0.15*30=4.5 | 0.2*30=6 | 0.65*30=19.5 |
| **Vacuum cleaner penetration** | (0.6*4.5) * 0.3 +(0.2*4.5) * 1.0 | (0.5*6) * 0.2 +(0.1*6) * 1.0 | (0.01*19.5) * (1.0) |
| **Total** | 1.71 | 1.2 | 0.195 |

**Total Vacuum cleaners in India could be= 3.105 crore.**

This is how you can solve a professional industry related question.

5) **How many sweets an average school going kid will eat in a year?(structural approach)**

For this question we will follow the same steps like we did before in the previous example.

**Step 1: define some terms yourselves to analyse the question.**

Unlike the previous question this does not have any subjective terms. The formal definition of sweets could be **"a small shaped piece of edible item majorly made out of sugar".** This question could vary with where the person is and how they perceive the term sweet. It does not matter how you consider it, all they(interviewer) expect from you is the number of cases you bring out into the situation and how you are going to split them in a meaningful way.

**Step 2: consider an example for the study.**

I consider an average school going kid, for example I would consider myself when I went to school and bring out factors for this guesstimate from my past version.

Cases where I had sweets in my schooling on an average year could be broadly classified into these categories.

- Celebrations: This could include
    - Birthdays of my close friends that I attend.
    - Marriages or house warming ceremonies of my relatives or associates.

- Meeting relatives to get together (these are different from the celebrations like marriages we attend).
- Sweets made by grandmother on some religious festivals.
- Any good news that happens within the family that is worth buying sweets and celebrating.
- National holidays
  - The Independence Day
  - Republic Day
  - Children's day and few other important one

**Step3: Split the audience (here you split sweets like in the previous example we actually split the people who watched XXX film).**

I have 7/8 friends who invite me for their birthdays. I attend 5 birthdays and eat one sweet at each birthday.

In a year, almost ¼ th of the time people would not prefer to get married or wish to have any functions because of some astrological significance so with the other ¾ th of the time I assume we get around 20 invitations round the year. We attend almost 75% of them so it would be 0.75*20 =15 out of which I eat sweet 80 percent of the time so 0.8*15 = 12

These get together do not happen quite often; these are a bit outliers at least in this example. I Consider 3 get together where I actually eat a sweet.

I go to my grandmother's house 3 times a year. This is where my major sweet consumption comes from. She makes almost 50 sweets every time for me to take back home (I eat only 30 percent of them, the rest portion goes to my family). I stay there for a week whenever I go, I eat 2 sweets per day. The total count would be 3*(0.3*50) +3*7*2 = 87.

Good news that is worth buying sweets: these do not happen so often I will consider 2 sweets.

There are many national holidays but school provides sweets only for republic day, Independence Day, it also provides sweets on children's day but it is not a national holiday. In total 3 sweets.

Total count would be = 5+12+3+87+2+3 =112. I eat almost 112 sweets a year.

# Data analytics Case Studies

6) **What are case studies and what are the types of case studies?**

Case studies are effective ways of bringing out various insights and results from the given stack of data or situation in a scientific research method. The main purpose of case studies is to understand the business problem or other real-world problems within the boundary of a specific organization, environment or individual.

There are essentially 4 different case studies:

- **Illustrative:** We study a familiar case in order to understand the big picture.
- **Exploratory:** In this you are given a dataset and expected to find out answers to various research questions. These are kind of direct questions which don't expect any deep analytical insight.
- **Cumulative:** It studies and collects information about different places and different times and sees how these factors influence the problem.
- **Critical Instance:** These case studies are used to find out reasons for the problem and analyse its impact.

7) **Number of matches played by each team in the year 2019 ODI format?**

**Introduction:**

Cricket has always been the favourite game of many of us in India. The data we take here is related to cricket and the format is ODI (50 over innings). The below are some of the entry columns used to describe the ODI life of international cricket for the year 2019.

1) Team 1
2) Team 2
3) Winner
4) Margin (either in wickets or runs): margin is shown in wickets, if the first batting team wins then margin is defined in terms of runs. In the second case when the team who batted 2nd wins then margin is defined in wickets.
5) Host: (venue city)
6) Match date

**Data set:**

| Team 1 | Team 2 | Winner | Margin | Ground | Match Date |
|---|---|---|---|---|---|
| New Zealand | Pakistan | New Zealand | 61 runs | Wellington | 06-Jan-18 |
| New Zealand | Pakistan | New Zealand | 8 wickets | Nelson | 09-Jan-18 |
| U.A.E. | Ireland | Ireland | 4 wickets | ICCA Dubai | 11-Jan-18 |
| New Zealand | Pakistan | New Zealand | 183 runs | Dunedin | 13-Jan-18 |
| U.A.E. | Ireland | Ireland | 67 runs | ICCA Dubai | 13-Jan-18 |
| Australia | England | England | 5 wickets | Melbourne | 14-Jan-18 |
| Bangladesh | Zimbabwe | Bangladesh | 8 wickets | Dhaka | 15-Jan-18 |
| New Zealand | Pakistan | New Zealand | 5 wickets | Hamilton | 16-Jan-18 |
| Ireland | Scotland | Ireland | 6 wickets | ICCA Dubai | 16-Jan-18 |
| Sri Lanka | Zimbabwe | Zimbabwe | 12 runs | Dhaka | 17-Jan-18 |
| Ireland | Scotland | Ireland | 24 runs | ICCA Dubai | 18-Jan-18 |
| New Zealand | Pakistan | New Zealand | 15 runs | Wellington | 19-Jan-18 |
| Australia | England | England | 4 wickets | Brisbane | 19-Jan-18 |
| Bangladesh | Sri Lanka | Bangladesh | 163 runs | Dhaka | 19-Jan-18 |
| Australia | England | England | 16 runs | Sydney | 21-Jan-18 |
| U.A.E. | Scotland | Scotland | 31 runs | ICCA Dubai | 21-Jan-18 |

These are some of the top rows present in the dataset.

In total we have 128 samples, that is 128 international ODI matches that happened that year. 5 of the matches either had no result or a tie. We have a total of 18 teams in the dataset.

**Approach:**

Prepare a structured way of thinking. Take small steps that's when you can actually address the question and think of a possible solution.

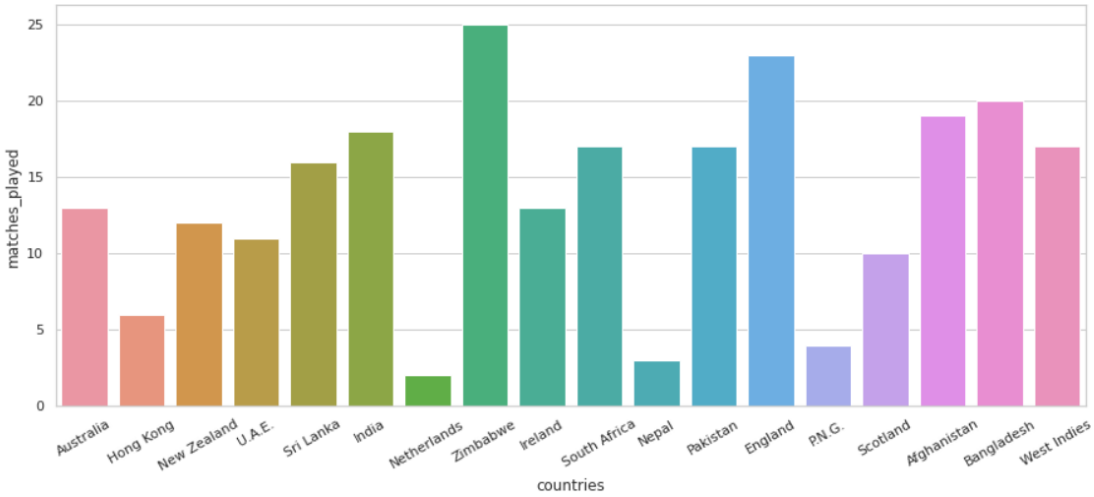
first, we have to find total number of countries actually played in 2019 ODI
next we think of number of matches played by each team

If you observe the dataset closely you find this is not a direct question, where you find the number of unique values in team1 or team 2 and get the total number of countries.
In fact, if you do that for team1 total unique values is 18 and team2 is 17 this happens because there is a chance that some teams are not mentioned in team1 column at all and some might be mentioned in both columns at different rows.

| | countries | matches_played |
|---|---|---|
| 0 | U.A.E. | 11 |
| 1 | Bangladesh | 20 |
| 2 | West Indies | 17 |
| 3 | Hong Kong | 6 |
| 4 | South Africa | 17 |
| 5 | Nepal | 3 |
| 6 | India | 18 |
| 7 | Pakistan | 17 |
| 8 | Sri Lanka | 16 |
| 9 | Zimbabwe | 25 |
| 10 | Australia | 13 |
| 11 | England | 23 |
| 12 | Ireland | 13 |
| 13 | New Zealand | 12 |
| 14 | Netherlands | 2 |
| 15 | P.N.G. | 4 |
| 16 | Scotland | 10 |
| 17 | Afghanistan | 19 |

**So how do we do it?**
We first find the unique values of both teams and apply union over them. This approach will give the total number of countries.

We are done with a major part of the question. Now, to find matches played by each team we just take each team name and search against both team1 and team2 and add them together (we know that a team cannot be in both team1 and team2 for the same row because a team cannot play with itself).




**8) Which team has the highest chance of winning if they bat first?**

In the dataset they did not mention which team has batted first. We can interpret it from the given data.

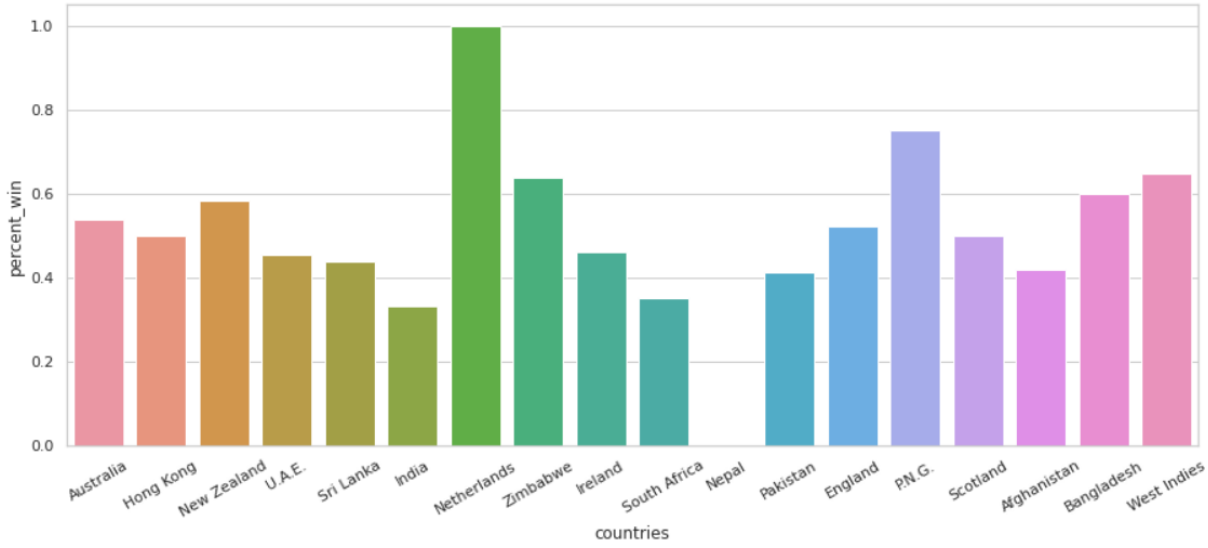
If the margin is in runs then the winning team batted first and if the margin is in wickets, then the winning team batted second.

To solve, we need to give some assumptions:

- The winning percentage with the first bat determines the true potential and advantage of a team.
- Number of matches played, even if it is extremely less, can determine the strength of a team.

Whenever you study a case study it is not the definite answer people are interested in. The interviewer wants to know your approach and metrics you consider in getting the solution and assumptions you took in order to get to it. It is important that you are aware of your assumptions when you solve a problem.

Tabular data of percentage win of each when first batted.




| | countries | percent_win |
|---|---|---|
| 0 | Australia | 0.538462 |
| 1 | Hong Kong | 0.500000 |
| 2 | New Zealand | 0.583333 |
| 3 | U.A.E. | 0.454545 |
| 4 | Sri Lanka | 0.437500 |
| 5 | India | 0.333333 |
| 6 | Netherlands | 1.000000 |
| 7 | Zimbabwe | 0.640000 |
| 8 | Ireland | 0.461538 |
| 9 | South Africa | 0.352941 |
| 10 | Nepal | NaN |
| 11 | Pakistan | 0.411765 |
| 12 | England | 0.521739 |
| 13 | P.N.G. | 0.750000 |
| 14 | Scotland | 0.500000 |
| 15 | Afghanistan | 0.421053 |
| 16 | Bangladesh | 0.600000 |
| 17 | West Indies | 0.647059 |

Netherlands has a high win percent, so theoretically this has the most edge over winning if it batted first. If you observe closely, you see that it only played 2 matches and you may say, it is unfair to come to that conclusion based on these 2 matches. The answer to this question is to go to assumption number 2.

Nepal in total played 3 ODI matches in 2019 and it never batted first, so we say we cannot find its chances of winning for this data set . India has the least chance of winning. It won 6 matches out of 18 matches it played first batting.

The main aim of these exercises is not to get definite answers but to give you an idea of how to approach similar things when you face it with huge real time datasets.

If you are not ok with having 2nd assumption. You can also come up with a metrics where you have certain kind of trade off between no of matches played and percent win (where a team has played high number of matches but low win percent but trade off brings it to top position)
Then you will assume that trade off metrics is best at representing the chance of a team batting first.

### 9) Which league team is ready to face hard game play situations next year?

**My assumptions:**

1) The team which plays at their own country has few advantages compared to the ones who play at other countries.
2) If a team has a record of playing more percent of matches under hard conditions that indicates that it is ready to face these conditions in future.

**Approach:**

considering these assumptions. Our approach would be to first map every city to its country and then iterate over each row and see which country played at other countries' venues and give a point to it. Doing this we end with a table having each country along with their score.

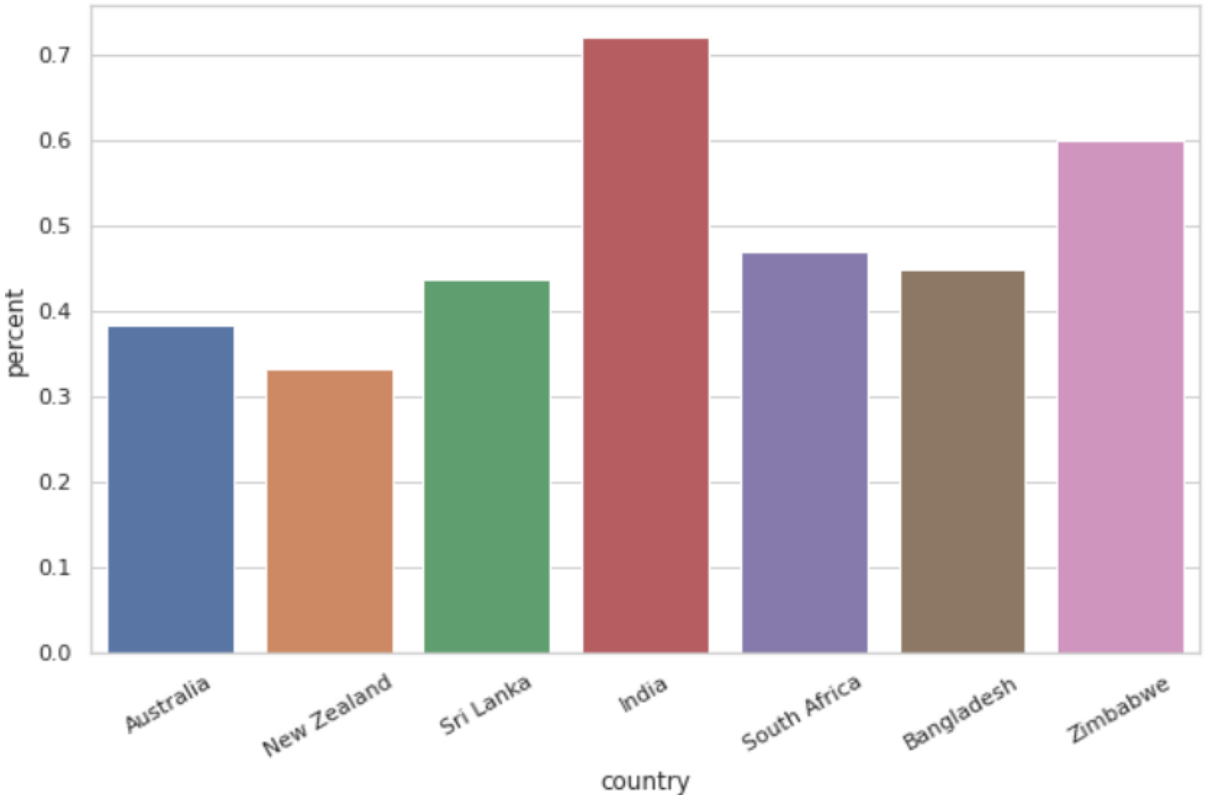
We now check the ratio of score over total matches played. The team with a high ratio is ready or adaptable to hard game play situations.

**Pointers to note:**

Sometimes each categorical variable needs special attention. In this example initially we have in total 18 countries but countries like P.N.G, Pakistan, Afghanistan,

Scotland, Hong Kong do not have home grounds even if they have they might not be recognized by ICC(international cricket council) to play ODI.

So if a country do not play at least a few matches in their home there is no point in comparing its strength at home versus foreign venue. to make things clear, I mentioned league teams in the question.




**Results and insights:** we observe India played more than 70% matches in foreign venues highest by any country so we can comfortably say that India can be pretty adaptable to foreign conditions in future and this is also the reason why India did not win many matches in 2019. Next, we have Zimbabwe followed by South Africa. New Zealand has the lowest percentage for playing at foreign venue.

# Case study 2

**Data set:**

The data set which we are going to study now is about Singapore residents.

The dataset runs from 2001 -2019. The government measures or laws implemented in a country need to consider all age groups and communities who might be affected by it. The other situation could be, suppose you are planning to establish an Indian restaurant in Singapore at a large scale. You should be aware of the growth of Indian community in Singapore over the years. You need to answer various questions such as

- What are the target age groups for your business?
- Is the Indian community active there?
- What is the population of other communities who might be interested in your business as customers?
    - Trend of Indian community over the years?
- Which age group Indian community is rising in Singapore?
  Knowing these you can make a lot of changes in the food menu and the way the interior decoration or the whole presentation of the restaurant.

This is the advantage of having a population dataset specific to ethnic composition and age group over the timeline (in years).

| | year | level_1 | level_2 | value |
|---|---|---|---|---|
| **0** | 2001 | Total Residents | 0 - 4 Years | 222991 |
| **1** | 2001 | Total Residents | 5 - 9 Years | 254077 |
| **2** | 2001 | Total Residents | 10 - 14 Years | 246972 |
| **3** | 2001 | Total Residents | 15 - 19 Years | 210903 |
| **4** | 2001 | Total Residents | 20 - 24 Years | 214636 |
| **...** | ... | ... | ... | ... |
| **6835** | 2019 | Other Ethnic Groups (Females) | 70 Years & Over | 2197 |
| **6836** | 2019 | Other Ethnic Groups (Females) | 75 Years & Over | 1348 |
| **6837** | 2019 | Other Ethnic Groups (Females) | 80 Years & Over | 858 |
| **6838** | 2019 | Other Ethnic Groups (Females) | 85 Years & Over | 454 |
| **6839** | 2019 | Other Ethnic Groups (Females) | 90 Years & Over | 190 |

6840 rows × 4 columns

The columns present in the dataset are:

- Year: Ranging from 2001 to 2019
- Level1: This is dividing population based on gender, ethnicity
    - Total Residents
        - Male residents
        - Female residents
        - Other ethnic communities
            - Male
            - female
        - Total Indians
            - Total Male Indians
            - Total Female Indians
        - Total Chinese
            - Total Male Chinese
            - Total Female Chinese
        - Total Malays
            - Total Male Malays
            - Total Female Malays

All of these are further filtered using level2

- Level2: dividing population based on age groups
    - 0-4 years
    - 5-9 years
    - 10-14 years
    - 15-19 years
    - 20-25 years
    - 25-29 years
    - 30-34 years
    - 35-39 years
    - 40-44 years,45-49 years,50-54years,55-59years,
    - Till 85-89 years
    - 65 and over,70 and over, 75 and over till 90 and over

We have a large number of age group filters so we will divide them based on our business.

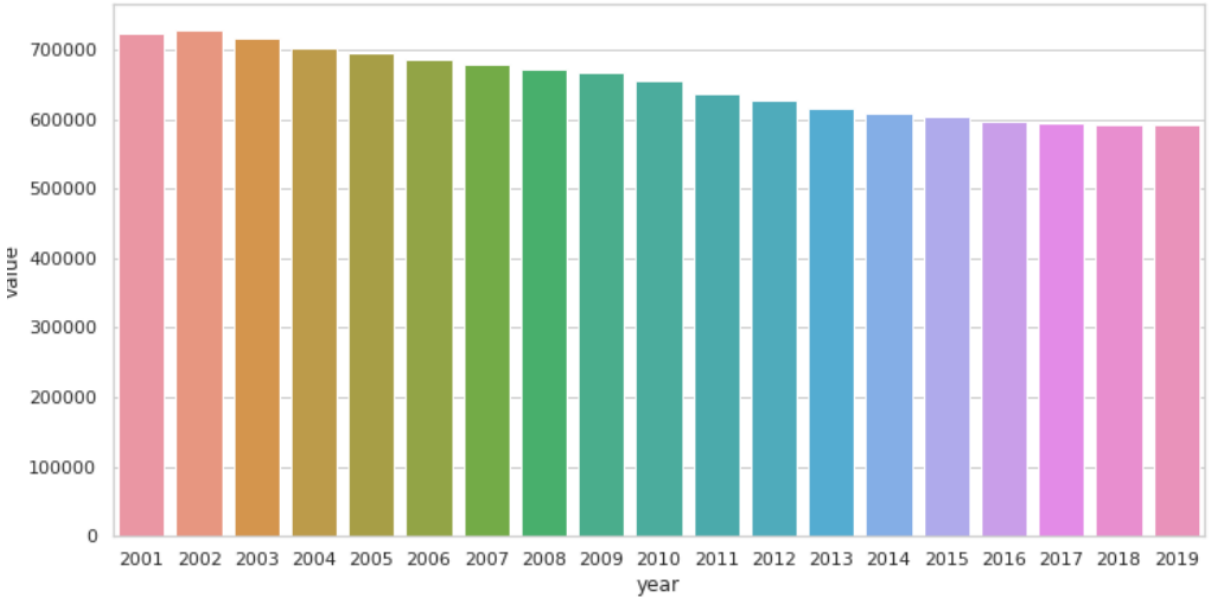
There are in total 6840 samples 19(years)*15(gender and community) *24(age based) = 6840

**10) Suppose I wanted to set up a cloth store for kids in Singapore, give me some suggestions as to what kind of clothing store I should set up and whom should I target?**

Firstly, let's merge some of the age groups that suit this business idea. The idea presented is clothing so generally we have two types of clothing based on age for kids 0-15, adults, senior citizens. It can also be based on gender male and female.

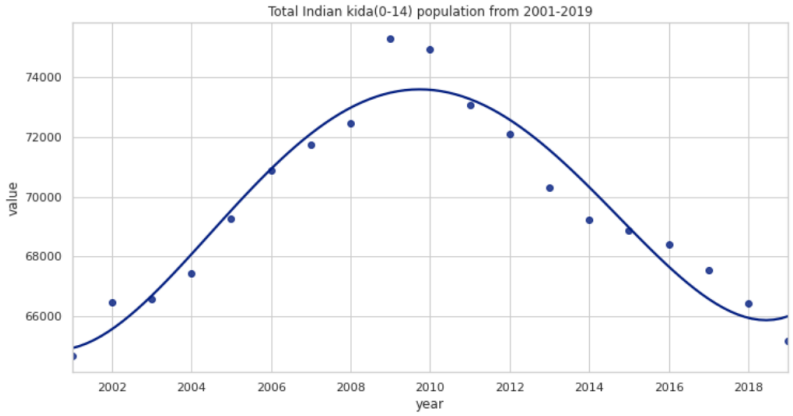
This question is open to interpretation. You can work on this question alone for days.

Firstly, I will see the trend of the strength of kids in Singapore.



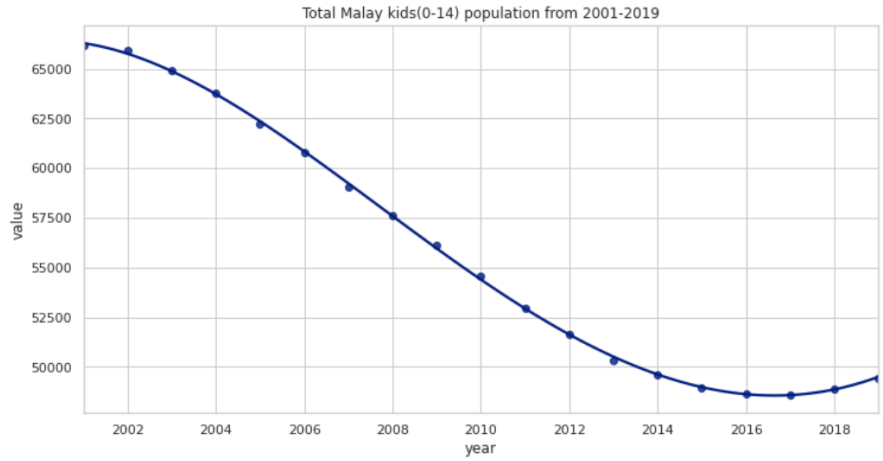

The total population of this group (age 1-15) is constantly decreasing.

Even India.




The total Indian kids data is a bit different from rest. This kind of has a maximum at 2010 and then it decreases but the $5^{th}$ degree polynomial

Adaption to these data points sees a possible increase in the number. If someone plans to set up a children clothing store for Indians it is not a bad idea.




Total Malay children population had a constant decrease till 2016 and from then on it took on an incremental path for the next 3 years so clothing specific to them might also be an added advantage.

# Case Study 3

**Spotify:**

Have you ever thought or wished that someone understands you and gifts or surprises with your favourite thing. You might have not noticed it but it is happening to you every single day**.** The video suggestions you get or the Netflix series you see on your feed all were suggested or appeared to you because there were some similarities between the content you already saw and the content that was displayed to you with the intention that you might like them too.

Here we are going to see Spotify data and understand how the company uses some of its information and suggests songs based on it.

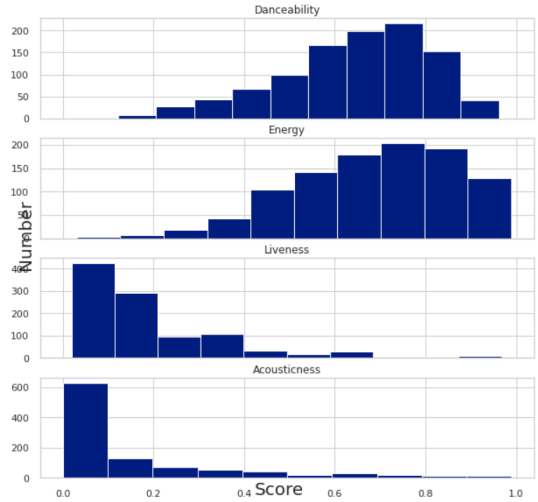
Let’s see a sample Spotify dataset below
The constraints of it are:

- acoustics
- danceability
- duration_ms
- energy
- instrumentalness
- key
- liveness
- loudness
- mode
- speechiness
- tempo
- time_signature
- valence target
- song_title
- artist

Most of the words are self-explanatory. Even if you do not understand some constraint completely think that it is some factor that helps or defines a music Like the loudness, duration etc.

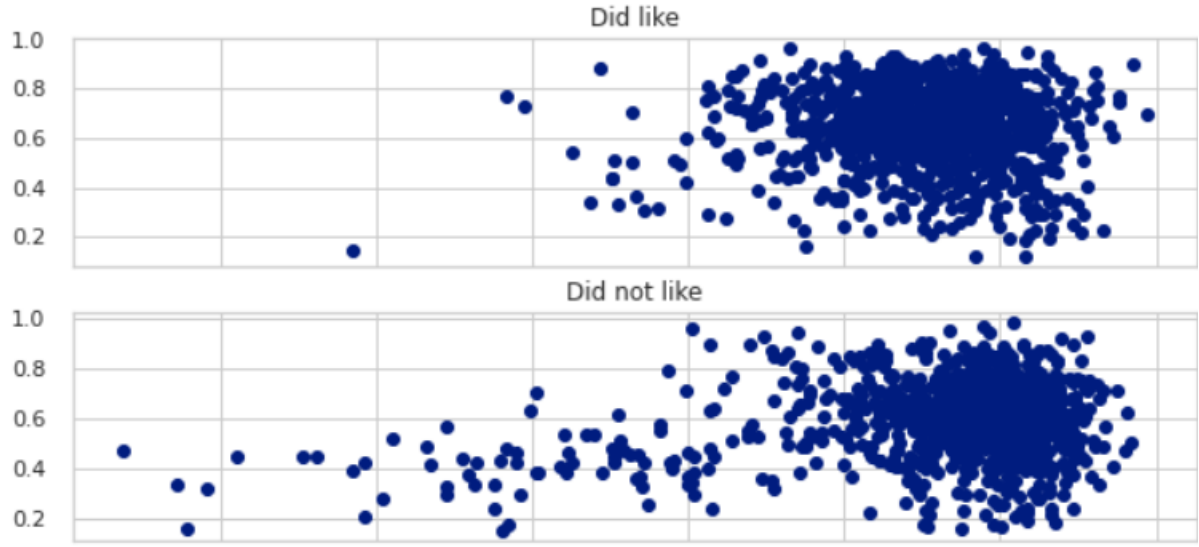
The target value tells if that particular person liked it or not: 1 if we liked it. 0 if we do not.
Now the question is

**11) based on the data set of a particular user suggest a song to the user**




Here I took 4 measures to see what is the taste of the songs which he liked (Target value =1)

We observed that this person likes his songs to have high danceability, energy, but low liveness, acoustics.




**X-axis loudness**
**Y-axis danceability**

**Observation:** we can set the Spotify user can like or dislike a song if it has high loudness and medium or average danceability but if the user had less loudness and average danceability then he is not going to like it probably see the 2nd Graph there are lots of samples in the left side when compared with the 1st Graph. So when both loudness and danceability are high then we cannot decide his taste solely basing these 2 constraints then we add more constraints like beats per second or energy and see how are the results for both the cases (one when he liked and other when he did not like it).
The question asked is a subjective question. You can work on this data alone for a number of days if you want to look at the actual data set it is here.

# Case Study 4

**12) Nike (footwear manufacturing company) has seen a decline of 20% in sales? Analyse this Case Study**

Let's follow some basic steps or procedures to approach this kind of case studies.

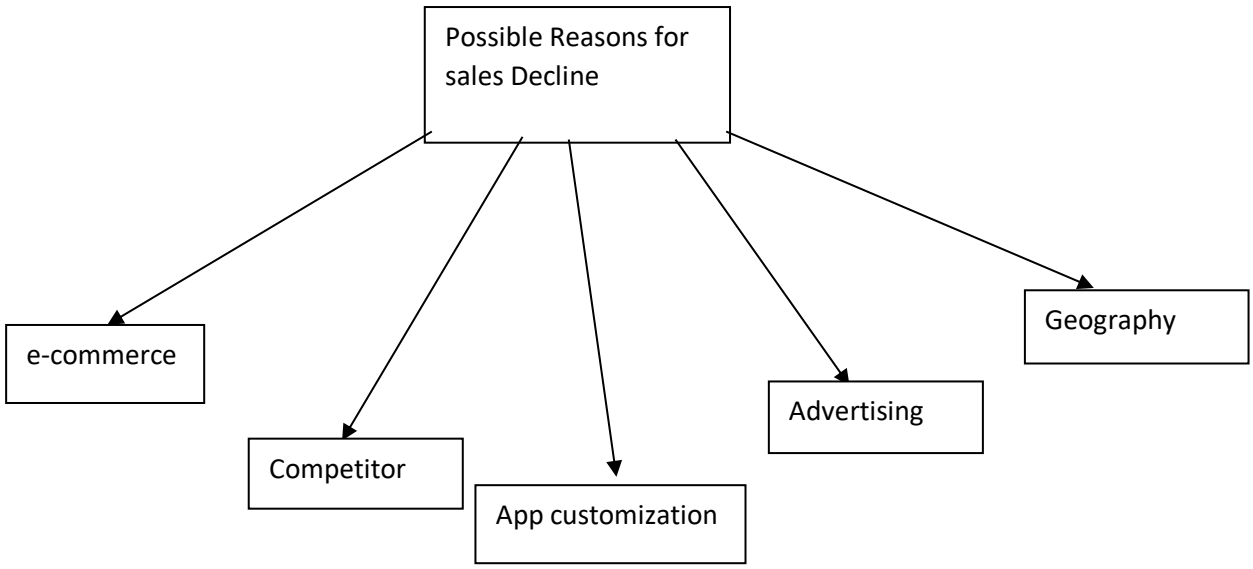
**Step1: Be clear of the terms and make sure you are with the same page as interviewer**

In the question we were told about decline in sales so know about what exactly is decline in terms of sales and how is it measured (like the time period). After you make sure you are clear with these things, go for the next step.

**Step-2: factors that affect your results that are asked in the case study**

In this case we need to find the reasons or possible factors for decline in Nike sales. First, know if this decline is a random thing or some outlier which is huge like the sales rate can never be constant, they go up and down for hundreds of reasons, first ask the interviewer what is the scale of this decline. Don't hesitate to ask a question because these things narrow things down further and make this case study easy for you.

Then come to analysing the factors. Some of the factors that I found would be worth analysing are given below.




I. **E-commerce:** There are various sources from where a customer can buy a Nike product. Some of the sources are
   - Nike own website
   - Amazon, Flipkart etc

   Analyse where exactly the decline is. Is it from my own website or other e-commerce site? If you can see a huge difference you found a factor for the answer then go deep and see what exactly causes it. Suppose you found a decline of around 50 percent in amazon so find out what's happening in it. How does the amazon search engine favour Nike. Is amazon delivering the products on time. Is there any decline in customer ratings? These could be some questions you can work on if you think E-commerce is a reason for decline.

II. **Competitor:** Nike has some huge competitors like Adidas, Puma, Reebok, Fila etc.
These international competitors might not always be the problem. There might be some local competitors who are creating brand value, notice them, see how Nike can dominate them.

III. **Advertising:** advertising is an essential part of a company's growth. There are lots of sources for advertising like in YouTube, television, websites, etc. Look at data and see how people react to it.
**For example:** If a company advertises it in YouTube then you can have the data on how many people were watching the full advertisements, number of people actually visiting the link provided in the description below the ad. Now, in this case see if these numbers have gone down lately. If they are going down then find out the reason for it. Is it because of any ad changes or did the company increase or decrease the time for the advertisements etc.

IV. **App customization:**
Once I saw a post on Linkedin which said "a 3 seconds additional delay in accessing amazon site or its search engine might cost it's company billions of dollars". There are these unnoticed reasons, which might affect sales hugely.
Having a cart option on a different position in the mobile app or making it difficult for people to make payments through the app.
What is the decline in sales with respect to IOS customers?
These are some of the factors in app customization that seriously affects the sales.

V. **Geography:** Once, I am working on real time company data in a competition. Previously, I thought it kind of equally represents all over the world with some domination in its origin country. I was wrong 80 % percent of it's orders or income comes from just North America. This helped me a lot in understanding the company structure. Europe with 9 percent is second so I took out Europe and saw how the income generated there and found out it is decreasing there from the past 5 years but the overall income is on increase like never before. What I meant to say from this is sometimes splitting things into geographical

regions and considering each individually gives lots of information which otherwise would be hidden.

# CASE STUDY-5

**13) How did KFC become a popular franchise in the world?**

First, we need to know how KFC started in the first place?
It was mid 1950s when Harland Sanders was in his 60s and intensively trying to market his “Colonel sander’s Recipe Kentucky Fried Chicken” which was prepared with 11 herbs and shrubs. Tables turned in his favour by 1963 KFC became a franchise of 300 outlets across the country. In 1964 he sold his franchise for 2 million dollars to Jack Massey and John Young Brown who became president and chairman respectively. Not until late 1960s KFC spread across the globe, it started its ventures in Japan, Hong Kong, Britain.

How did these foreign establishments become successful?
It's because of the connections they had when they established KFC in foreign countries. For example: KFC signed a contract with Mitsubishi Shoji Kaisha Ltd. in Japan. Subsidiaries were also established in Great Britain, Hong Kong, South Africa, Australia, New Zealand, and Mexico. They did not just go there and establish themselves.

**Problems in Expanding**
Initially KFC found it difficult to expand because of cultural differences. For example: In Germany people are not habituated to take-out food or ordering food but KFC got its success quickly in Asian countries because chicken is a staple dish. In some countries there are some poultry regulations, example: Malaysia.

**How did it solve?**
The company has been most successful in foreign markets when local people operate the local establishments. The point here is to think like a local, not like an American company.

**14) Take example of one country and show how did KFC cross its hurdles to set up its operations in that country**

Let's consider India. KFC opened its first outlet in India, Bangalore in 1995. KFC was aware of economic liberalization in the 1990s and managed to get permission for 30 outlets.

**Lesson:** Law of land plays a very important role in establishing business on foreign land. KFC did fail sometimes in adapting to it, like we discussed in the above question, even in India it hasn't been that easy, KFC was initially welcomed with several protests from farmers, environmentalists, Doctors.

**Initially hardships when established**

Apart from the protests there had been many hardships for KFC to succeed in India. Till 1998 KFC did not pick up in its sales so they went closely to look into the matter.

Some of the things they found are:

- KFC was targeted for the upper middle class and rich but their strength in terms of populations is not high enough to support KFC operations.
- People felt that KFC was only a place for chicken. Indian mobs always loved to have diverse options in their menu.
- KFC was also believed to be expensive.
- KFC was initially positioned not as a family restaurant but as a teenage hangout place so Indians who prefer family space did not opt to go to KFC.

**How did KFC overcome these issues?**

- KFC altered its traditional menu to suit Indian nativity. Coleslaw, a "New England salad" , was replaced with green fresh salads. During the Navratri festival, KFC offered a new range of nine vegetarian products, which included Paneer burgers. Added family combos for both vegetarian and non-vegetarians.
- KFC also changed its positioning. Now its message seeks to attract families which includes some recreation in addition to food.
- Kids fun corner is a place for children to play within the KFC establishment. Games like ball pool and Chicky Express have been

introduced for kids. The company also introduced a meal for kids at Rs. 60, which was served with a free gift.

These are few among many changes KFC did over the years to attract Indians to its franchise.