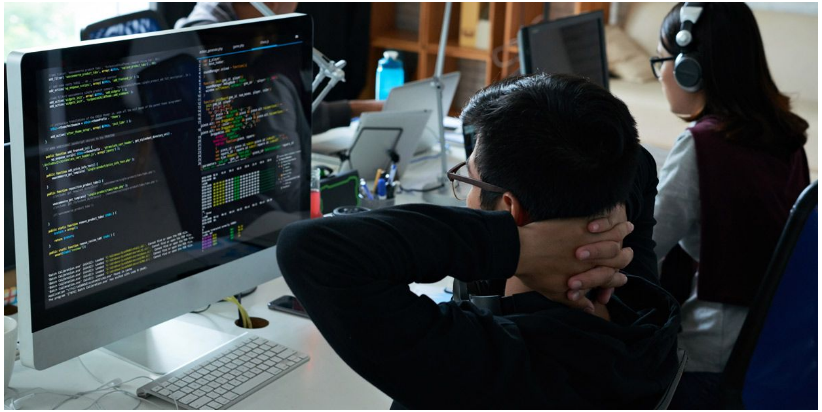
# Probability for Data Science

Analytics Vidhya
Learn everything about analytics

# Relevance of Probability

- A Data scientist without the knowledge of Probability and Statistics is like a Pilot without the knowledge of aerodynamics.



# Relevance of Probability

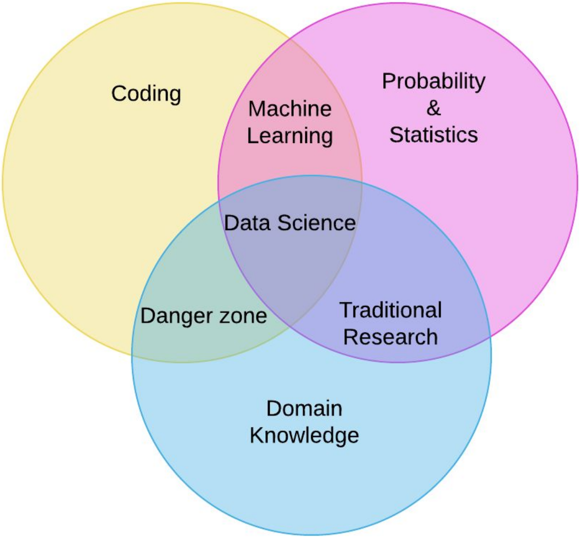
- Improve Business Sense

Ex. Bank providing better services to the customers who are likely to churn.

Analytics Vidhya
Learn everything about analytics

# Relevance of Probability

- Foundational language of Data Science




Analytics Vidhya
Learn everything about analytics

# Definition of Probability

- Probability is the ***likelihood*** or ***chance*** of an **event** occuring.

**5-20%**

Probability of selling to a **new prospect**

**60-70%**

Probability of selling to an **existing customer**

Analytics Vidhya
Learn everything about analytics

# Churn Prediction

Problem : Identify customer who will churn?

| | gender | age | occupation | churn |
|---|---|---|---|---|
| 0 | Male | young | salaried | 0 |
| 1 | Male | young | self_employed | 0 |
| 2 | Male | old | self_employed | 0 |
| 3 | Male | young | self_employed | 0 |
| 4 | Female | young | salaried | 1 |
| 5 | Male | old | salaried | 0 |
| 6 | Female | young | self_employed | 1 |
| 7 | Male | young | self_employed | 0 |
| 8 | Male | young | salaried | 1 |
| 9 | Male | young | salaried | 0 |
| 10 | Male | young | self_employed | 1 |
| 11 | Female | young | self_employed | 1 |
| 12 | Male | young | retired | 0 |
| 13 | Female | young | self_employed | 0 |
| 14 | Male | old | self_employed | 0 |

Analytics Vidhya
Learn everything about analytics

# Random Experiment

*Random experiment* is a process with a number of possible outcomes.

Those outcomes are not necessarily certain.

Ex. The profession of a customer?

Self Employed

Salaried

Retired

Analytics Vidhya
Learn everything about analytics

# Sample Space

*Sample Space associated* with a random experiment is a set of all possible outcomes.

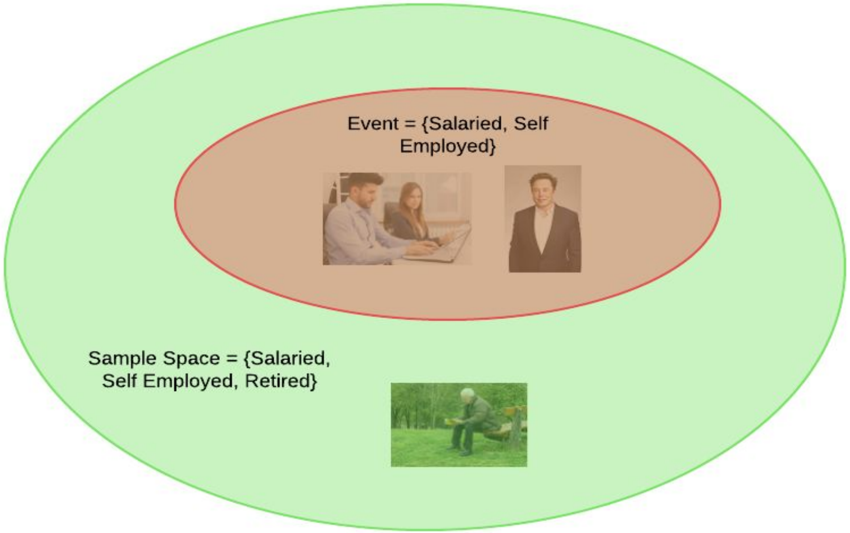
Ex. The profession of a customer?


Analytics Vidhya
Learn everything about analytics

# Event

*An event* is a subset of Sample Space.

Ex. If customer is currently working ?

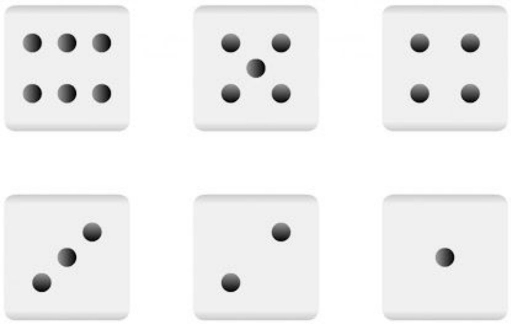



Analytics Vidhya
Learn everything about analytics

# Probability

The probability of an outcome O is a number P, between 0 and 1 that measures the likelihood that O will occur.

P = 0 → Impossible outcome.
Ex. Getting 8 on rolling a six faced dice



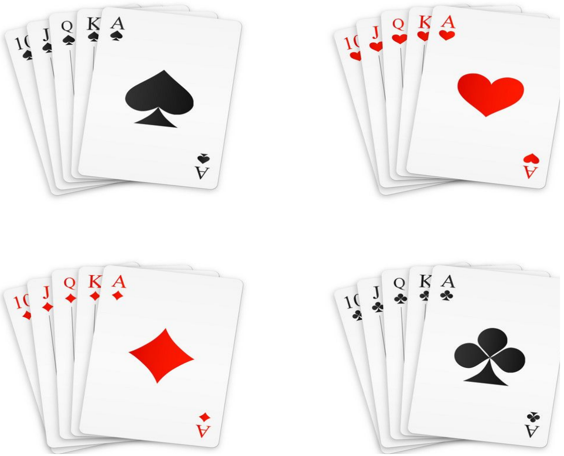
Analytics Vidhya
Learn everything about analytics

# Probability

The probability of an outcome O is a number P, between 0 and 1 that measures the likelihood that O will occur.

P = 1 → Definite outcome

Ex. Winning a pot with royal flush

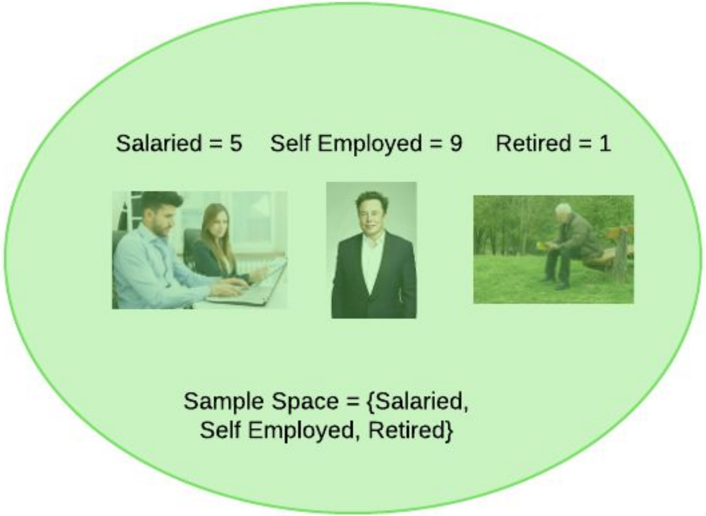


Analytics Vidhya

Learn everything about analytics

# Probability of a random experiment

Ex. The profession of a customer?




P(Customer = Salaried) = 5/15 = 0.333
P(Customer = Self Employed) = 9/15 = 0.6
P(Customer = Retired) = 1/15 = 0.067

Analytics Vidhya
Learn everything about analytics

# Definite and Impossible outcome in our data

Ex. If a salaried woman is going to churn ?

Analytics Vidhya
Learn everything about analytics

# Definite and Impossible outcome in our data

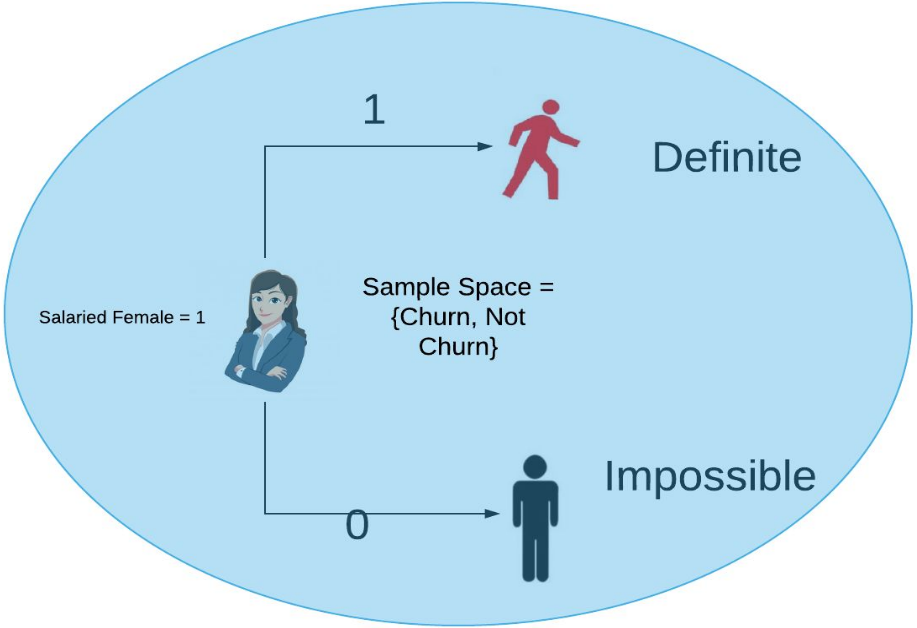
Ex. If a salaried woman is going to churn ?

Churn = 1

Churn = 0

# Definite and Impossible outcome in our data

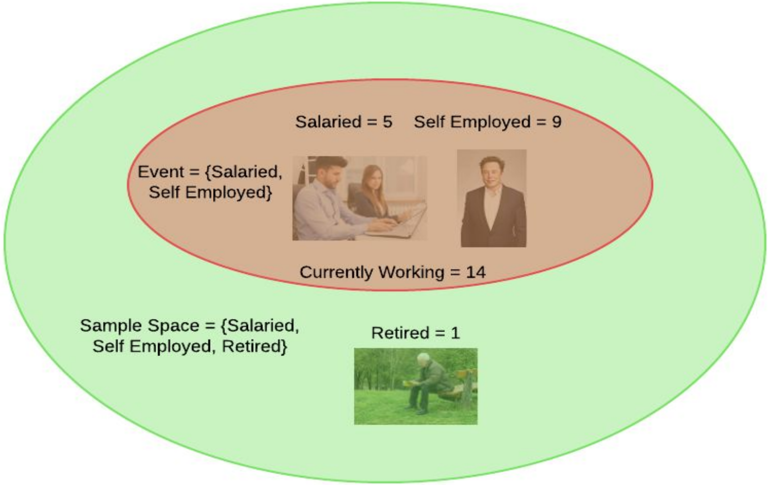
Ex. If a salaried woman is going to churn ?




Analytics Vidhya
Learn everything about analytics

# Probability of an Event

Ex. If a customer is currently working ?




P(customer = working) = 14/15 = 0.933

Analytics Vidhya
Learn everything about analytics

# Axioms of Probability

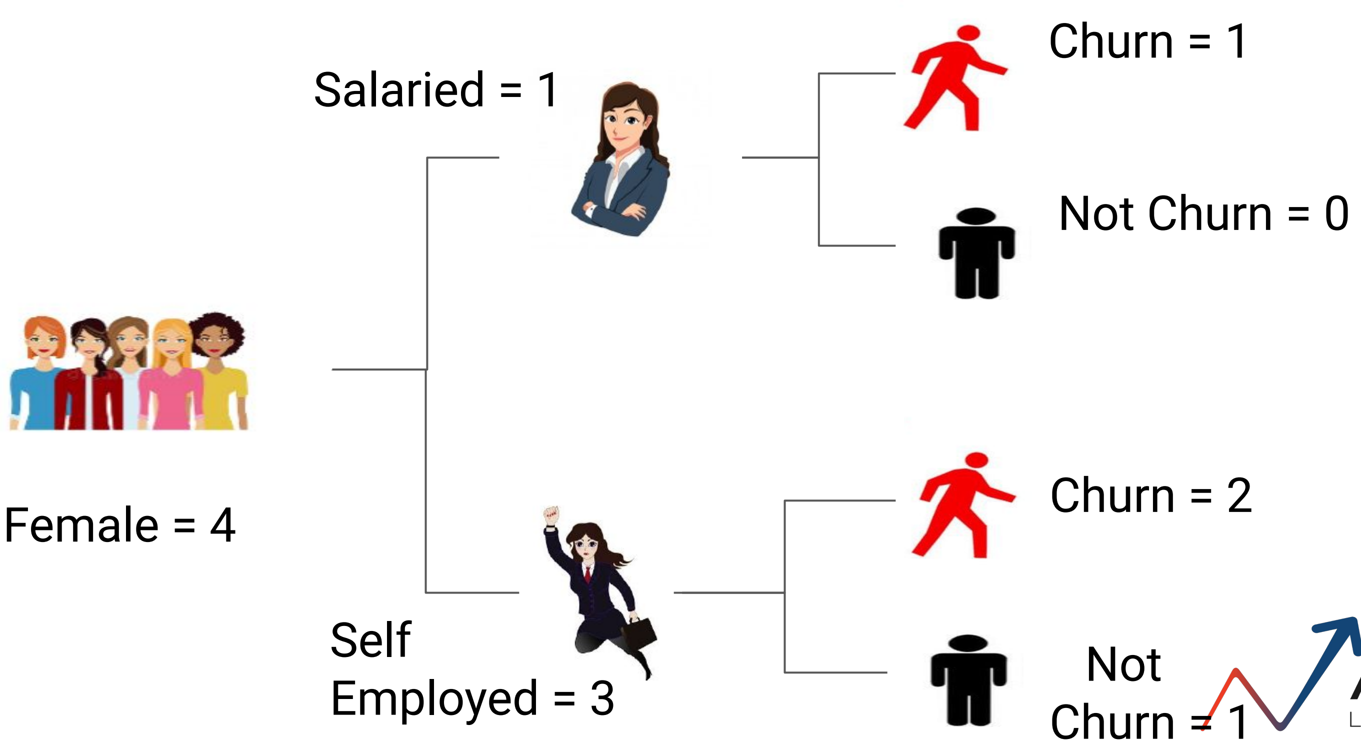
- For all events E, $0 \leq P(E) \leq 1$
- For Sample Space, $P(S) = 1$
- $P(A \cup B) = P(A) + P(B)$ for mutually disjoint events. ( $A \cap B = 0$)

  Mutually disjoint (can not both occur at once), mutually exclusive, have no elements in common.

Analytics Vidhya
Learn everything about analytics

# Axioms of Probability

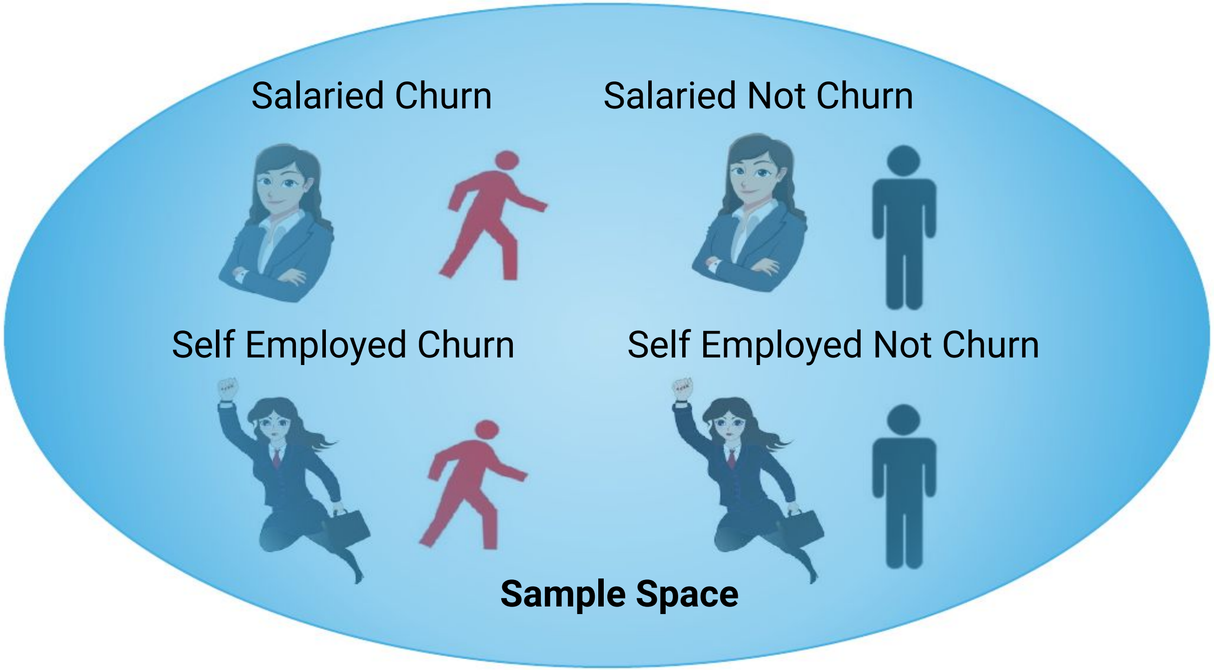
Ex. Profession of a female customer and her chances that she is going to churn ?




Analytics Vidhya
Learn everything about analytics

# Axioms of Probability

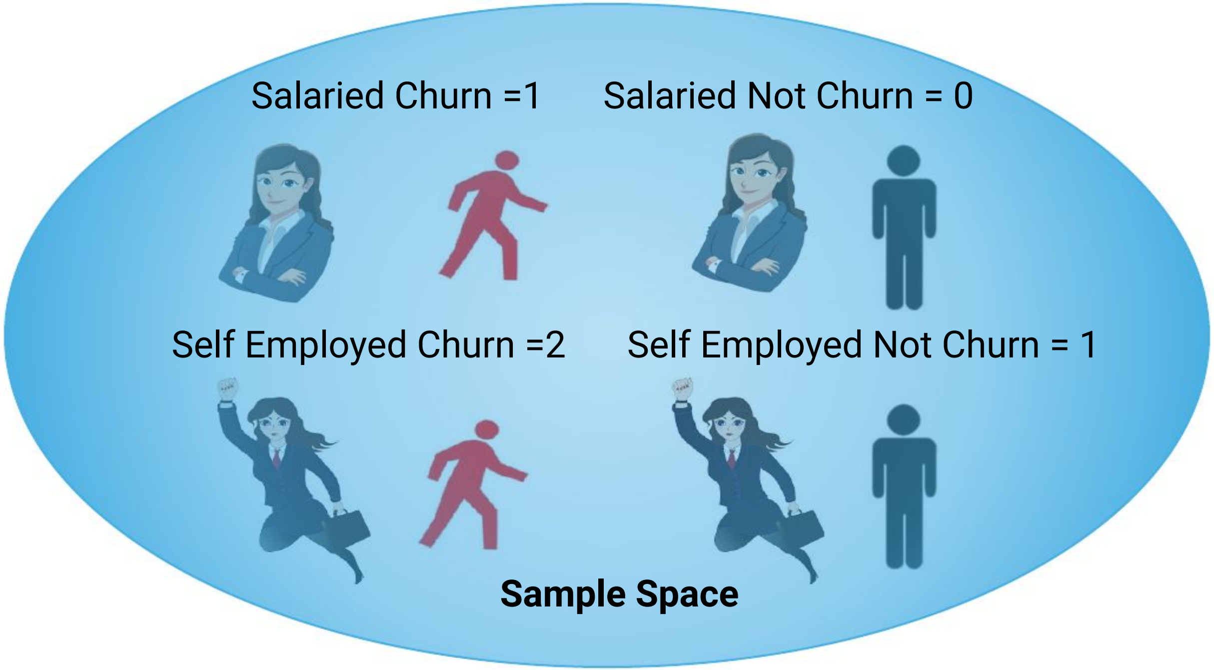
Ex. Profession of a female customer and her chances that she is going to churn ?




Analytics Vidhya
Learn everything about analytics

# Axioms of Probability

Ex. Profession of a female customer and her chances that she is going to churn ?




P(Salaried Churn) = $\frac{1}{4} \rightarrow 0.25$
P(Salaried Not Churn) =$0 \rightarrow 0$
P(Self Employed Churn)= $\frac{1}{2} \rightarrow 0.5$
P(Salaried Churn) = $\frac{1}{4} \rightarrow 0.25$

P(Sample Space)
$= 0.25 + 0 + 0.5 + .25 = 1$

Analytics Vidhya
Learn everything about analytics

# Complement

The complement of an an event E in a sample space S , denoted $E^c$, is the collection of all outcomes in S that are not elements of the set E.

Event E : getting an even number on rolling dice.

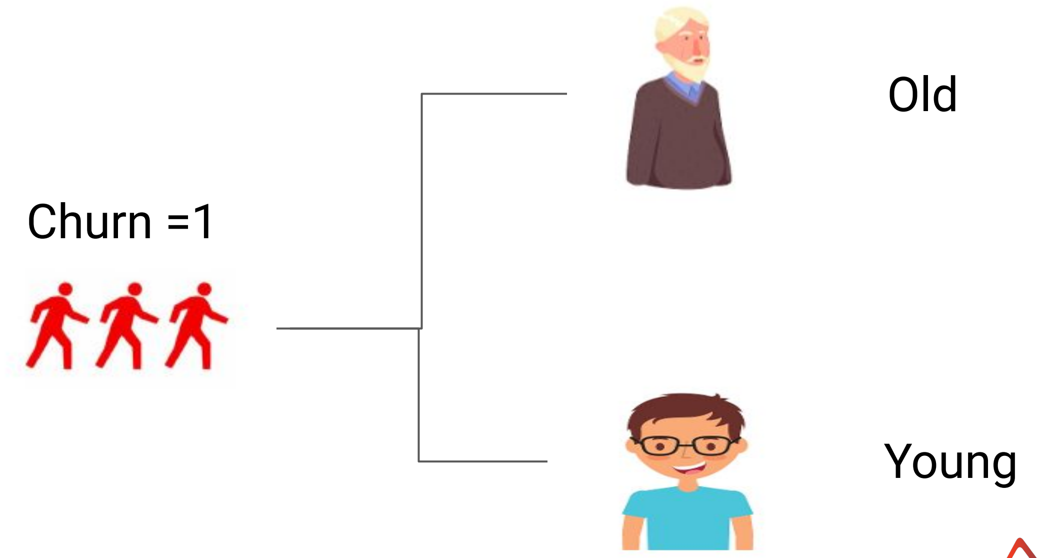
Event $E^{c}$: getting an odd number on rolling dice.

The Probability rule for Complements $P( E^c ) = 1 - P(E)$

Analytics Vidhya
Learn everything about analytics

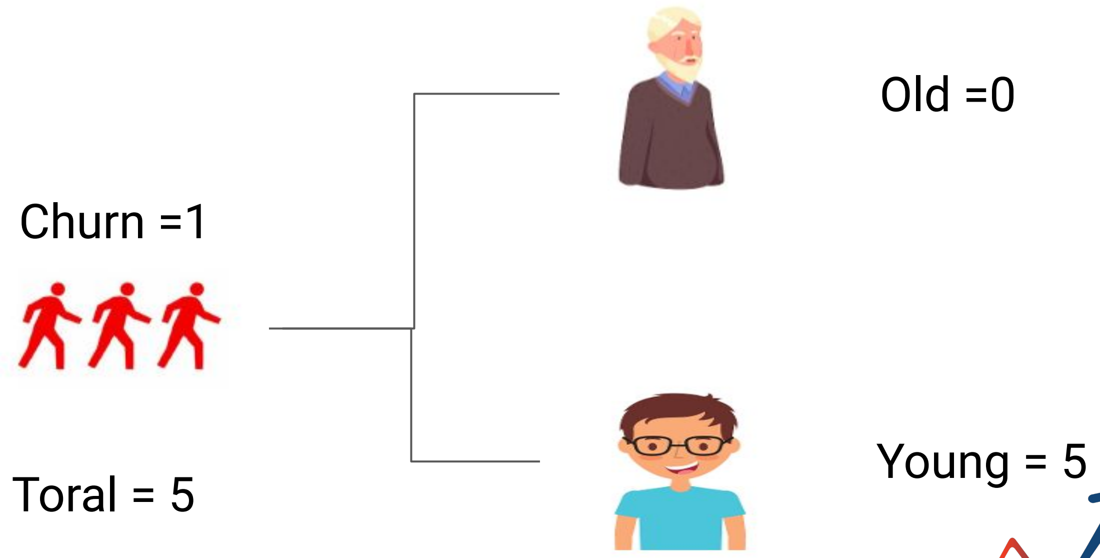
# Complement

Ex.  Among all the customers who are going to churn, Probability of one being an old customer ?



# Complement

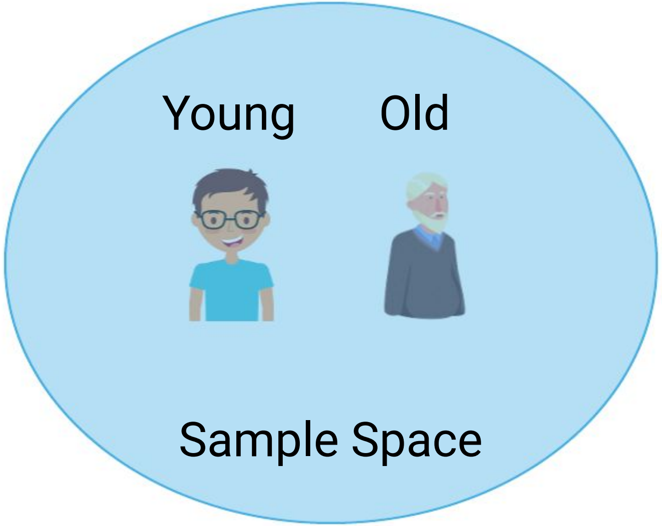
Ex. Among all the customers who are going to churn, Probability of one being an old customer ?




Analytics Vidhya
Learn everything about analytics

# Complement

Ex.  Among all the customers who are going to churn, Probability of one being an old customer ?

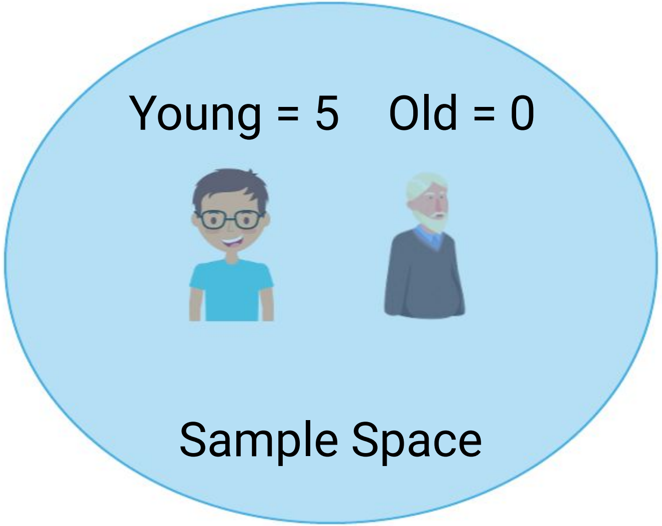



Analytics Vidhya
Learn everything about analytics

# Complement

Ex. Among all the customers who are going to churn, Probability of one being an old customer ?

E : Old
$E^c$ : Young




$P(E) = 1 - P(E^c)$

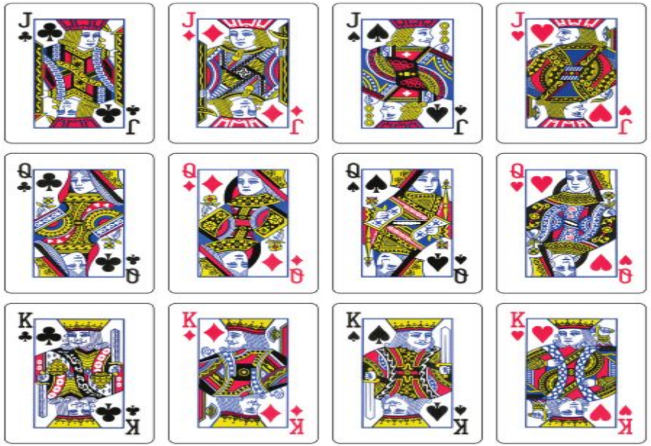
$= 1 - 1$

$= 0$

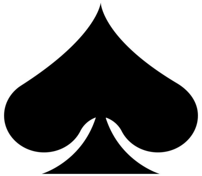
# Intersection

Denoted by $\cap$, is the collection of all outcomes that are common in events .

Event A : getting a face card.



Event B : getting a spade card.



Analytics Vidhya
Learn everything about analytics

# Intersection

Denoted by $\cap$, is the collection of all outcomes that are common in events .

Intersection of A and B: Card being a face card of spade?


Analytics Vidhya
Learn everything about analytics

# Intersection

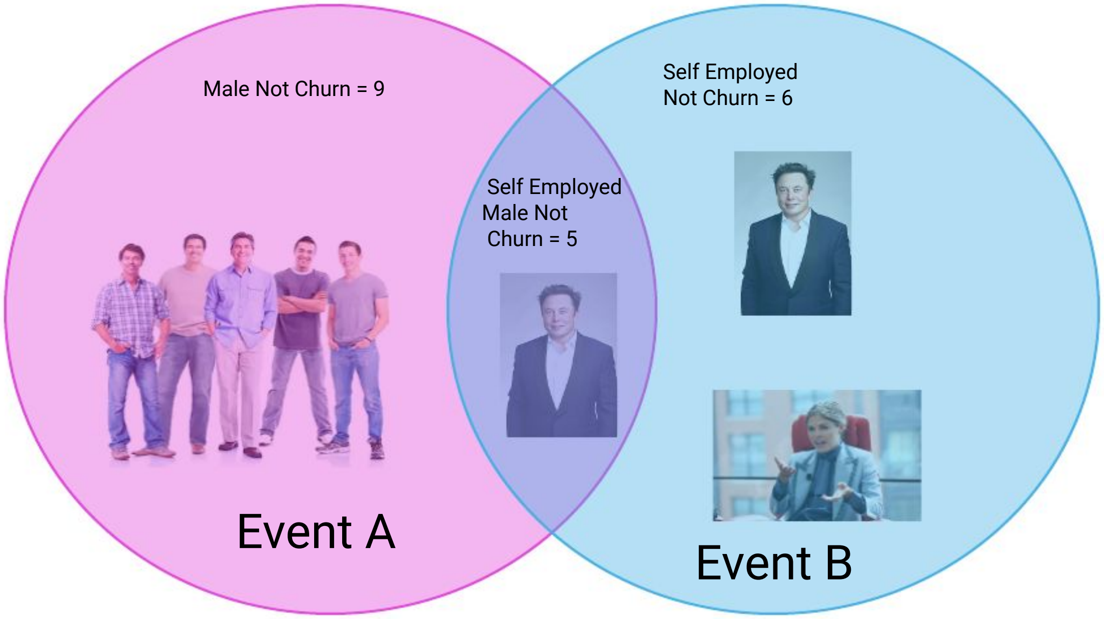
Denoted by $\cap$ , is the collection of all outcomes that are common in events .

Ex: Among all the customers who are not going to churn, Probability of one being a Self Employed Male person ?

Analytics Vidhya
Learn everything about analytics

# Intersection

Ex: Among all the customers who are not going to churn, Probability of one being a Self Employed Male person ?




Total Retaining Customers = 10

$P(A \cap B) = 5/10 \rightarrow 0.5$

Analytics Vidhya
Learn everything about analytics

# Union

Denoted by $\cup$**,** is the collection of all outcomes that are elements of **any** of the events .

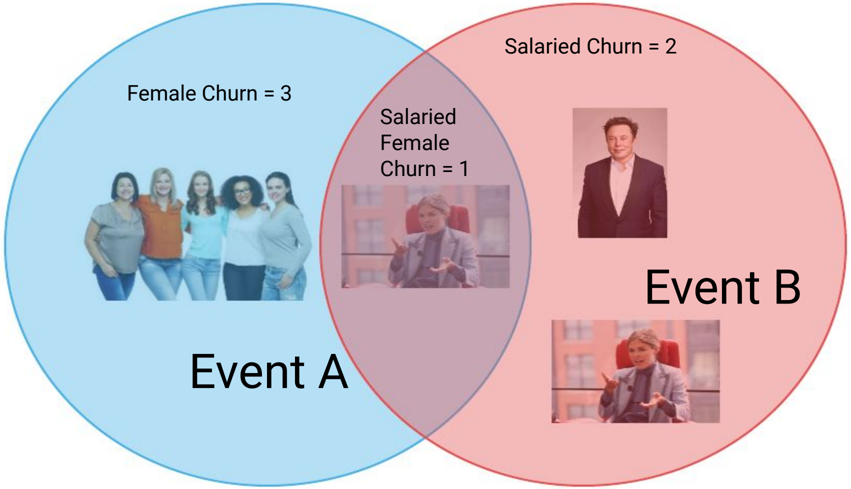
Ex: Among all the customers not going to churn, One being a Female or a Self employed person?

Analytics Vidhya
Learn everything about analytics

# Union

Denoted by $\cup$, is the collection of all outcomes that are elements of **any** of the events .

Ex: Among all the customers not going to churn, One being a Female or a Self employed person?




Total Customers reducing interaction = 5

$P(A \cup B) = 4/5 \rightarrow 0.8$

Analytics Vidhya
Learn everything about analytics

# Mutually Exclusive Event

Mutually Exclusive Events : Two events, A and B are said to be mutually exclusive if they can not occur together (have no common elements)

$P(A \cap B) = 0$

Event A : Getting a number greater than 4 rolling a dice.

** Different from compliment

Event B : Getting a number less than 3 rolling a dice.

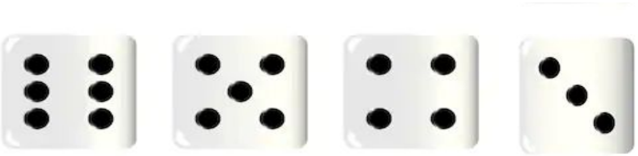
Analytics Vidhya
Learn everything about analytics

# Mutually Exhaustive Event

Mutually Exhaustive Events : Mutually exhaustive means that the events together make up everything that can possibly happen

$$P(\cup E^{n}_{i=1}) = S$$

Event A : Getting a number greater than 3 rolling a dice.



** Different from Mutually Exclusive

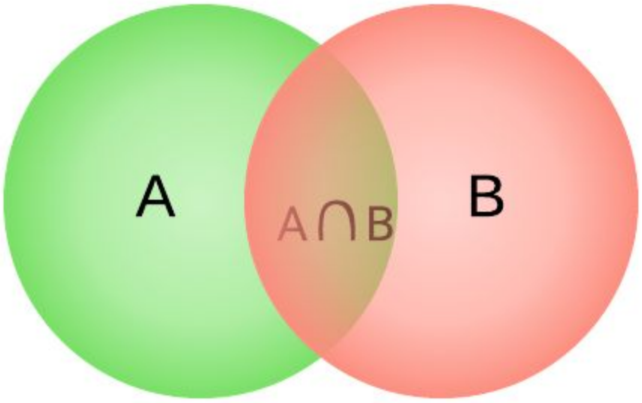
Event B : Getting a number less than 4 rolling a dice

Analytics Vidhya
Learn everything about analytics

# Additive Rule

Additive rule of Probability :-

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$




Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

The conditional Probability of A given B, denoted P(A | B) , is the Probability that event A has occurred given that event B has *definitely* occurred.

$$P(A \mid B) = P(A \cap B) / P(B)$$

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

Ex. Probability of a Customer is going to churn given that she is a female.

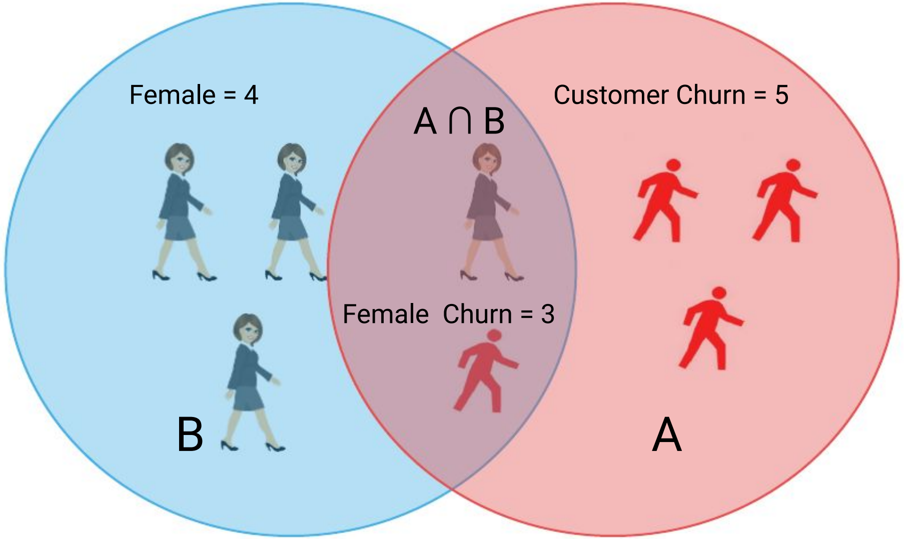
Event B = Customer being a female

Event A = Customer Churn

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

Ex. Probability of a Customer is going to churn given that she is a female.




Total Customers → 15

$$P(A \mid B) = P(A \cap B) / P(B)$$

$$P(A \mid B) = (3/15) / (4/15)$$

$$= 3/4 \rightarrow 0.75$$

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

What if $P(A \mid B) = P(A)$ ??

Event B has no impact on the likelihood of Event A.

A is **independent** of the event B.

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

For two independent events A and B, $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$

By Conditional Probability Calculations $P(A \mid B) = P(A \cap B) / P(B) = P(A)$

For two events to be independent $P(A \cap B) = P(A) . P(B)$

Analytics Vidhya
Learn everything about analytics

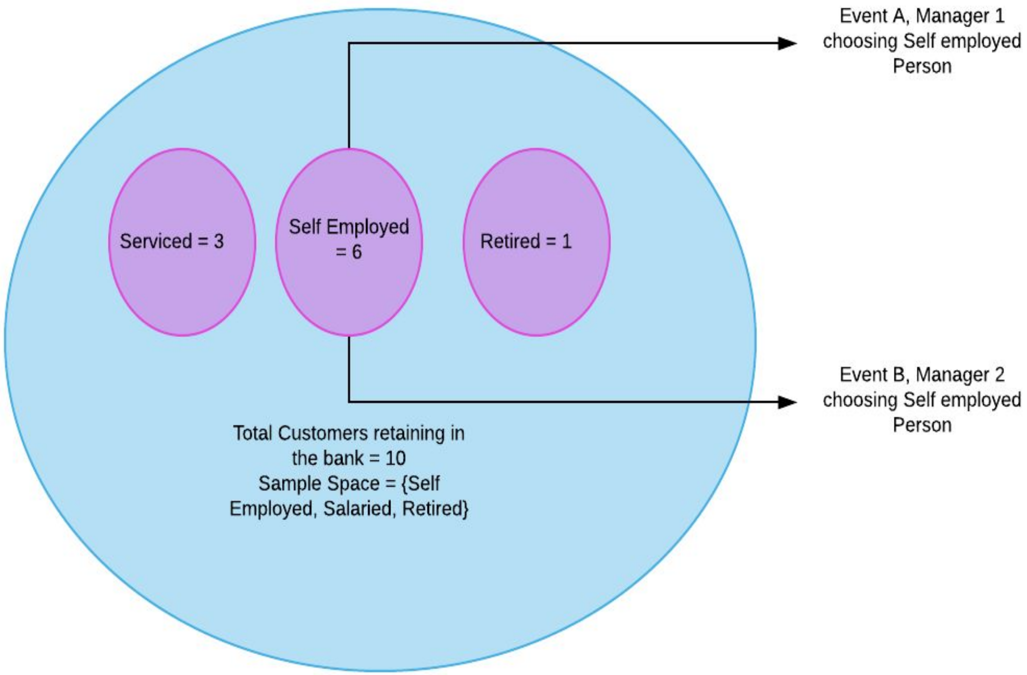
# Conditional Probability and Dependent & Independent Events

Ex. The bank has given the task to 2 different branch managers (who don't communicate with each other) to offer new loan scheme to one existing customer.

They both *randomly* go to existing customers to offer the scheme.

What is the probability that both of them will choose self employed customer?

# Conditional Probability and Dependent & Independent Events




Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

What is the probability that both of them will choose self employed customer?

$$
\begin{aligned}
P(A \cap B) &= 6/10 \times 6/10 \\
&= 36/100 \\
&= 0.36
\end{aligned}
$$

Independent Events (Probability with replacements)

# Conditional Probability and Dependent & Independent Events

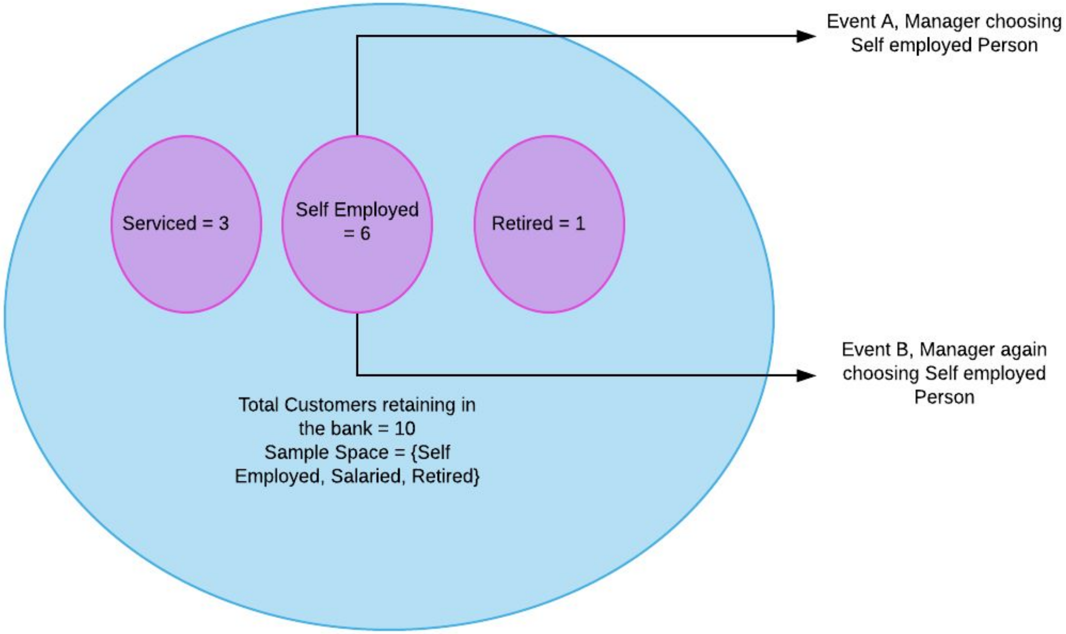
Ex. The bank has given the task to a branch managers to offer new loan scheme to two existing customers.

He *randomly* goes to existing customers to offer the scheme.

What is the probability that he will choose two self employed customer?

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events




Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

What is the probability that Manager will choose two self employed customer?

$$P(A \cap B) = 6/10 \times 5/9$$
$$= 30/90$$
$$= 0.333$$

Dependent Events (Probability without replacements)

Analytics Vidhya
Learn everything about analytics

# Conditional Probability and Dependent & Independent Events

An alternative to calculate the conditional probability

$$P(A \mid B) = P(A \cap B) \,/\, P(B) \quad \text{.... (1)}$$

$$P(B \mid A) = P(B \cap A) \,/\, P(A) \quad \text{.... (2)}$$

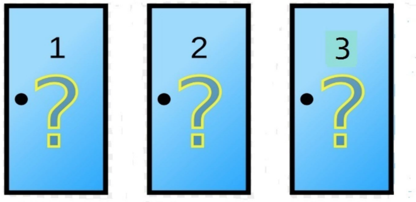
$$P(A \mid B) = P(B \mid A)\ .\ P(A) \,/\, P(B)$$

**Bayes Theorem**

Analytics Vidhya
Learn everything about analytics

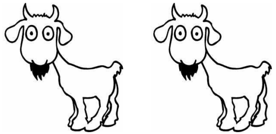
# Monty Hall Problem

Suppose you are in a game show and you are given three doors.




Behind one door is a car

Behind the others, goats.



Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

The host asks you to pick a door, He knows exactly what’s behind the doors

After your pick
He opens one such door which has goat behind it.

Now He asks you if you want to switch to the door he has not revealed yet or stick with your first choice ?

What’d you do ???

Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

Our Brain says it doesn't matter ?

Probability of both the doors (One you Picked and One not revealed by the host) will be equal for having car = 1/2

Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

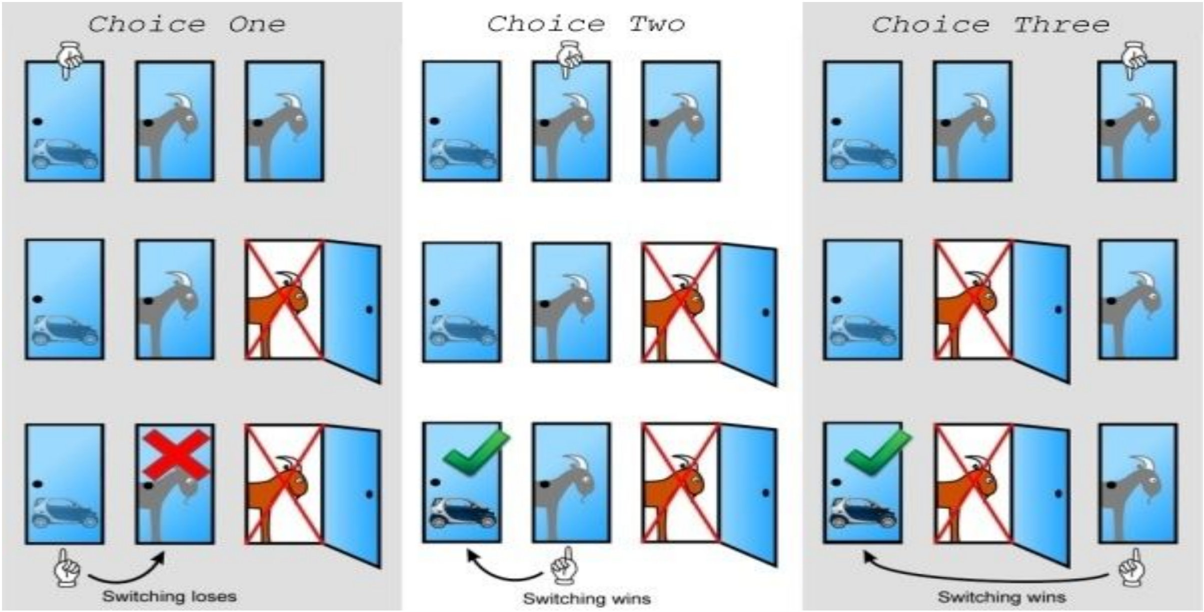
Let's think psychologically first,
If there would have been 10000 doors and host showed you all the door except one, would you switch then ?

Probably Yes.
Why ?

Self Realization

Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem




Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

More intuition???

Let's get back to the three doors problem.

If you don't switch You will win the car only if you were correct in guessing right door initially
Prob = ⅓ = 0.33

If you switch You will win the car only if you were not correct in guessing right door initially = ⅔ = 0.66

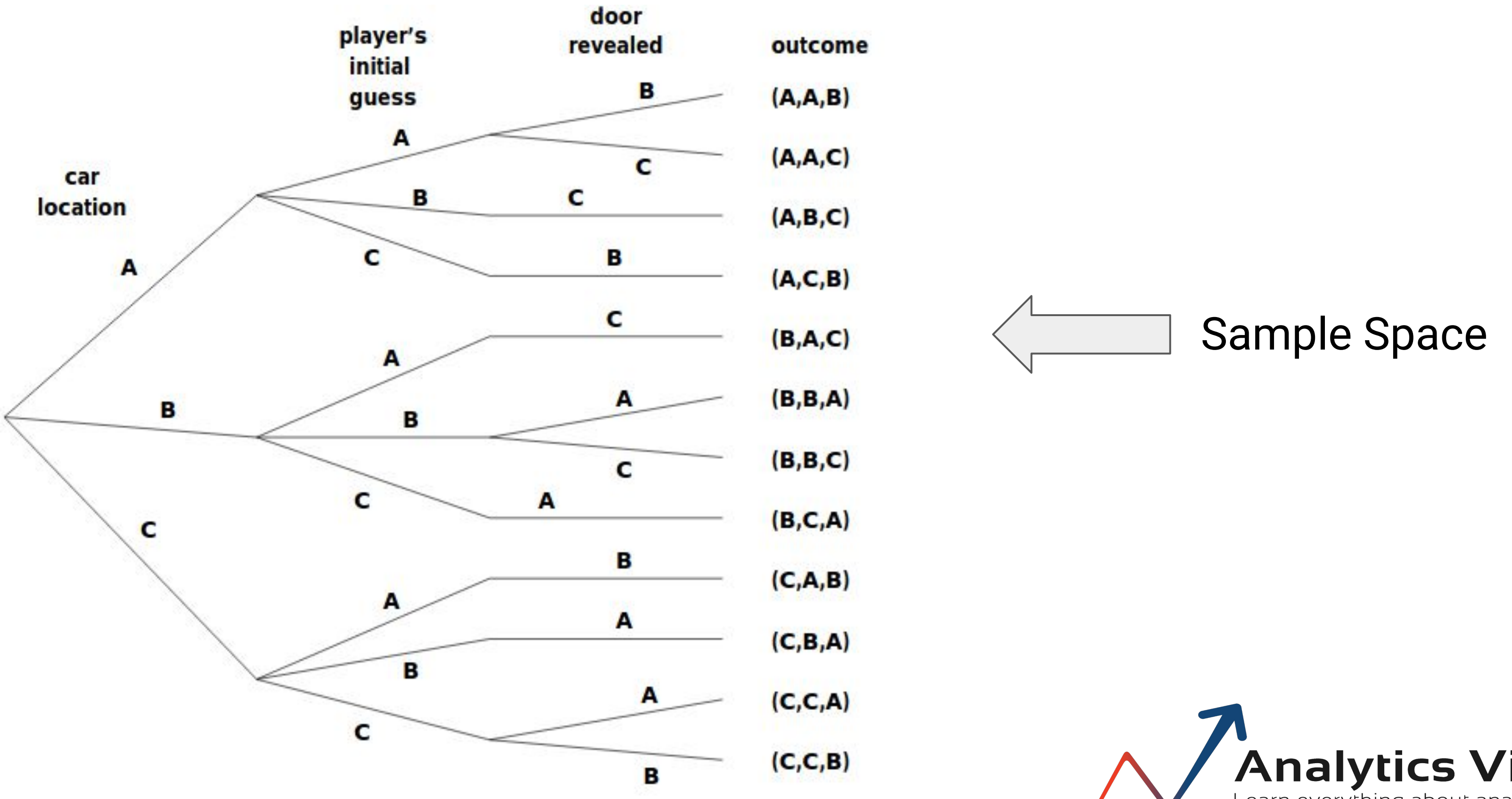
Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

Three randomly determined quantities here,

1. The door concealing the car.
2. The door player picked initially.
3. The door host opens to reveal.

# Monty Hall Problem




Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

Event A → The Prize is Behind C

({C,A,B}, {C,B,A}, {C,C,A},{C,C,B})

Event B → The player guessed correctly at first place

{A,A,B}, {A,A,C}, {B,B,A},
{B,B,C}, {C,C,A}, {C,C,B}

Analytics Vidhya
Learn everything about analytics

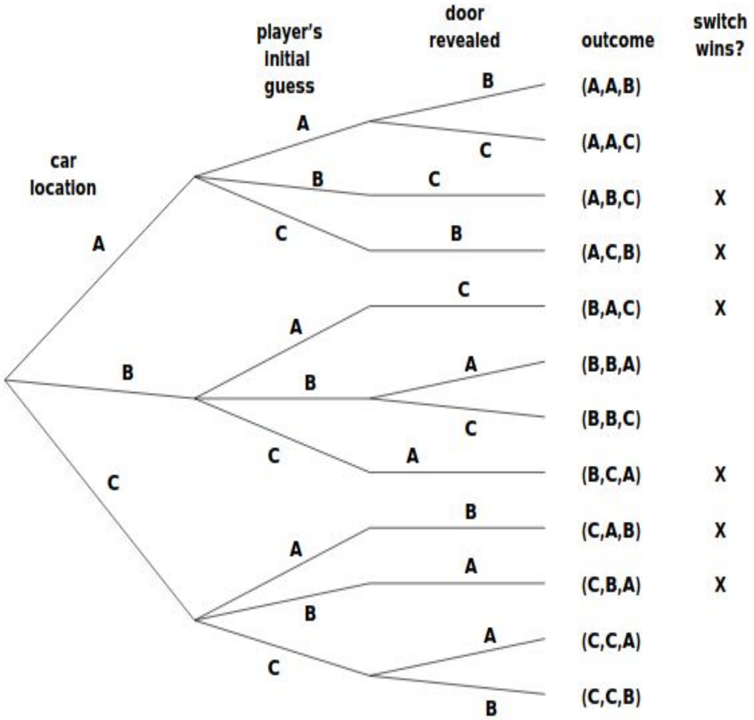
# Monty Hall Problem

Our Event → Player wins when switch ??

{A,B,C} {A,C,B} {B,A,C} {B,C,A} {C,A,B}, {C,B,A}

Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

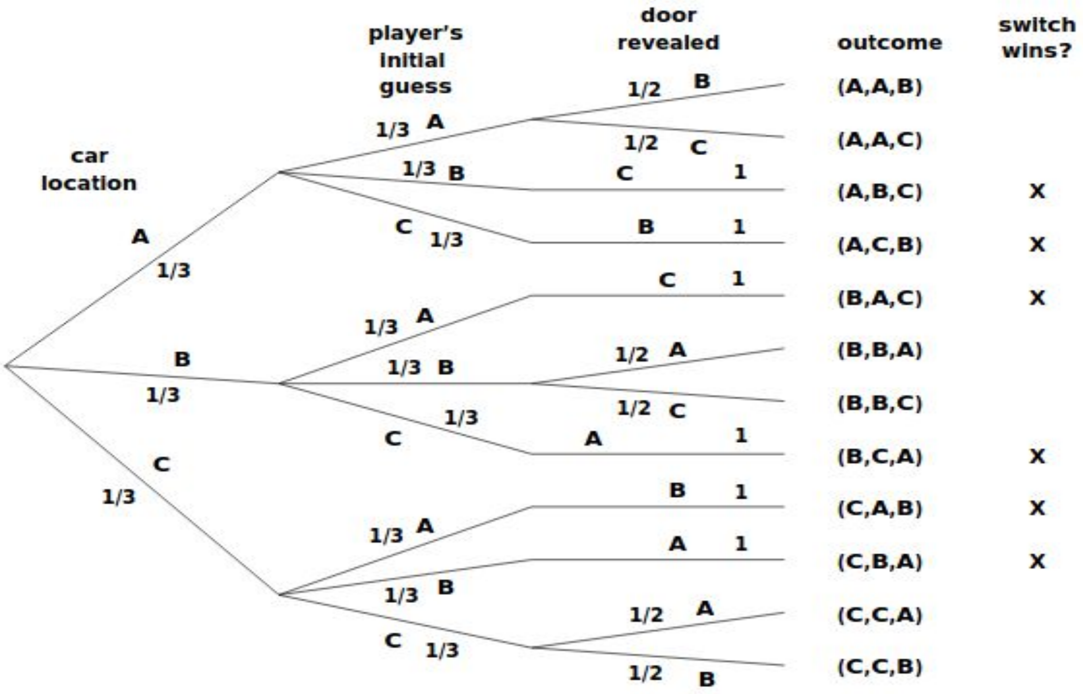



Winning in 6 Out of 12 cases so ½ = 0.5, Probability ??

No, The reason is that these outcomes are not all equally.

Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem




Analytics Vidhya
Learn everything about analytics

# Monty Hall Problem

Probability (Switching wins)

$= (P\{A,B,C\} + P\{A,C,B\} + P\{B,A,C\} + P\{B,C,A\} + P\{C,A,B\} + P\{C,B,A\})$

$= 1/9 + 1/9 + 1/9 + 1/9 +1/9 + 1/9$

$= 2/3$

Analytics Vidhya
Learn everything about analytics