SURVIVAL ANALYSIS

CANCER DATA

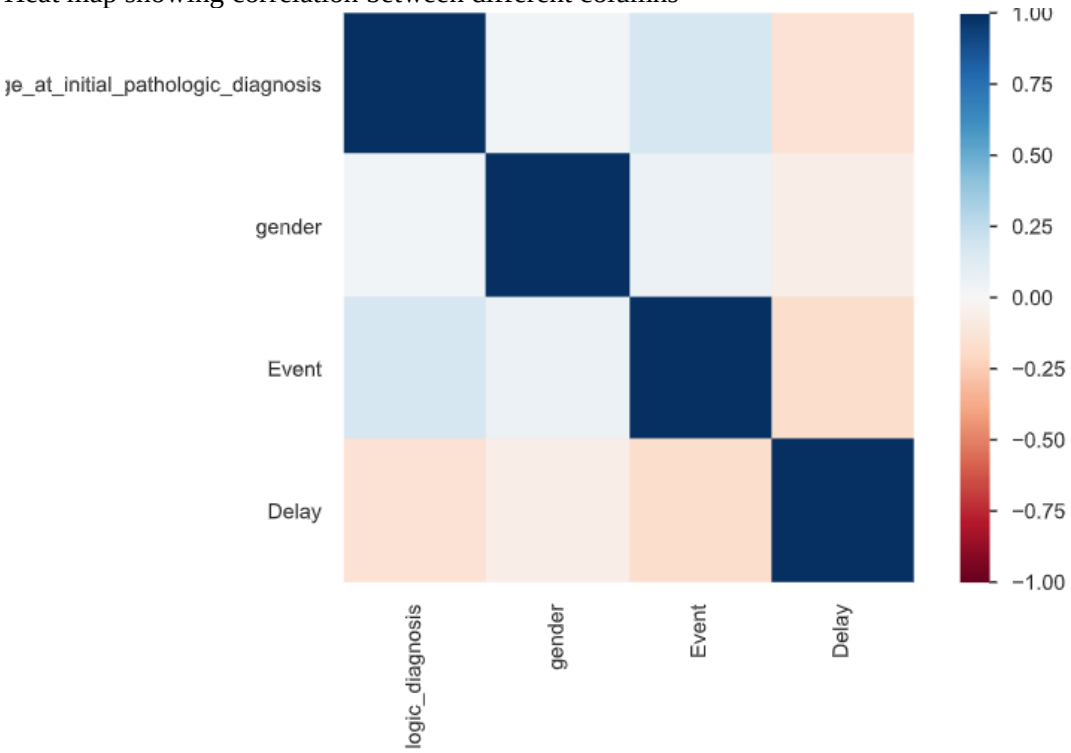
BY- SHIVANI

INTRODUCTION

Sex is known to be an important factor in pathogenesis, diagnosis and treatment of cancers, and it has been an independent prognostic factor for several cancer sites. Sometimes, races and stage of cancer are also important factor in treatment of cancer.
In this report I have done some basic analysis as well as survival analysis by relating some of the common factor that shows the survival chances of people of different races, different gender and have different cancer stage etc. The AJCC staging system is used to describe most types of cancer. Also called TNM staging system.
Data collected from different races i.e., Asians, White, American Indian, Black or African Americans and Native Hawaiian.

ANALYSIS

Heat map showing correlation between different columns

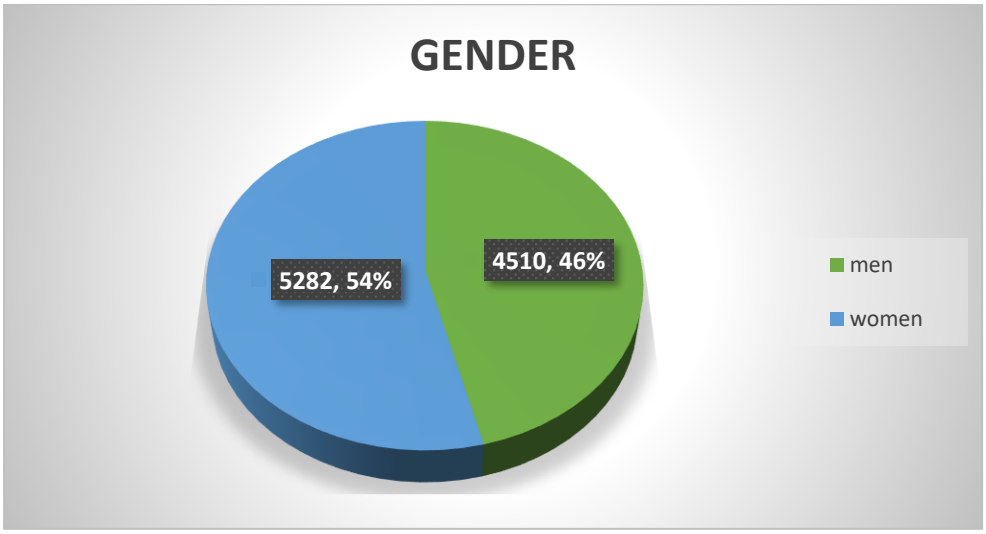


Total Number of Observation	9792
GENDER	
Male	4510
Female	5282
EVENT	
Survived (0)	6439
Not- survived (1)	3353

Table -1- Observation

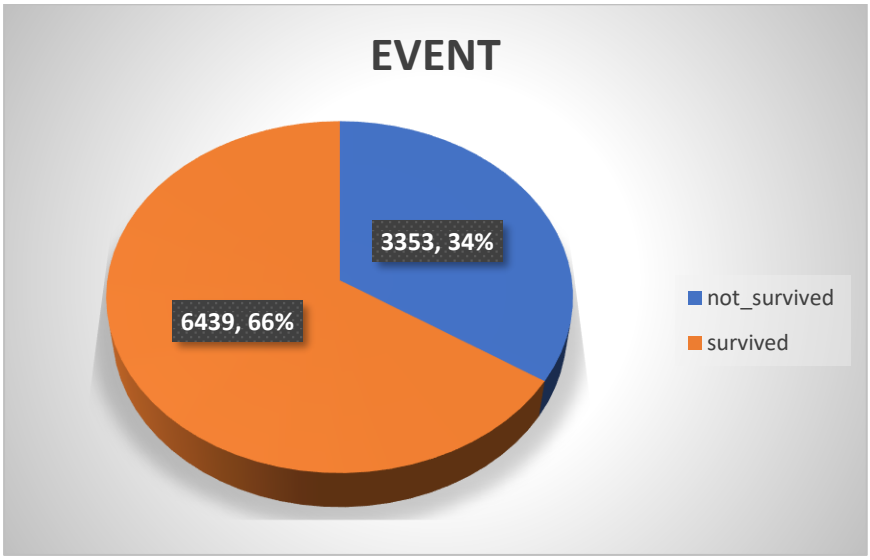
➔ **DISTRIBUTION OF GENDER**

Out of 100, 54% are females and 46% are male.



➔ **EVENT**

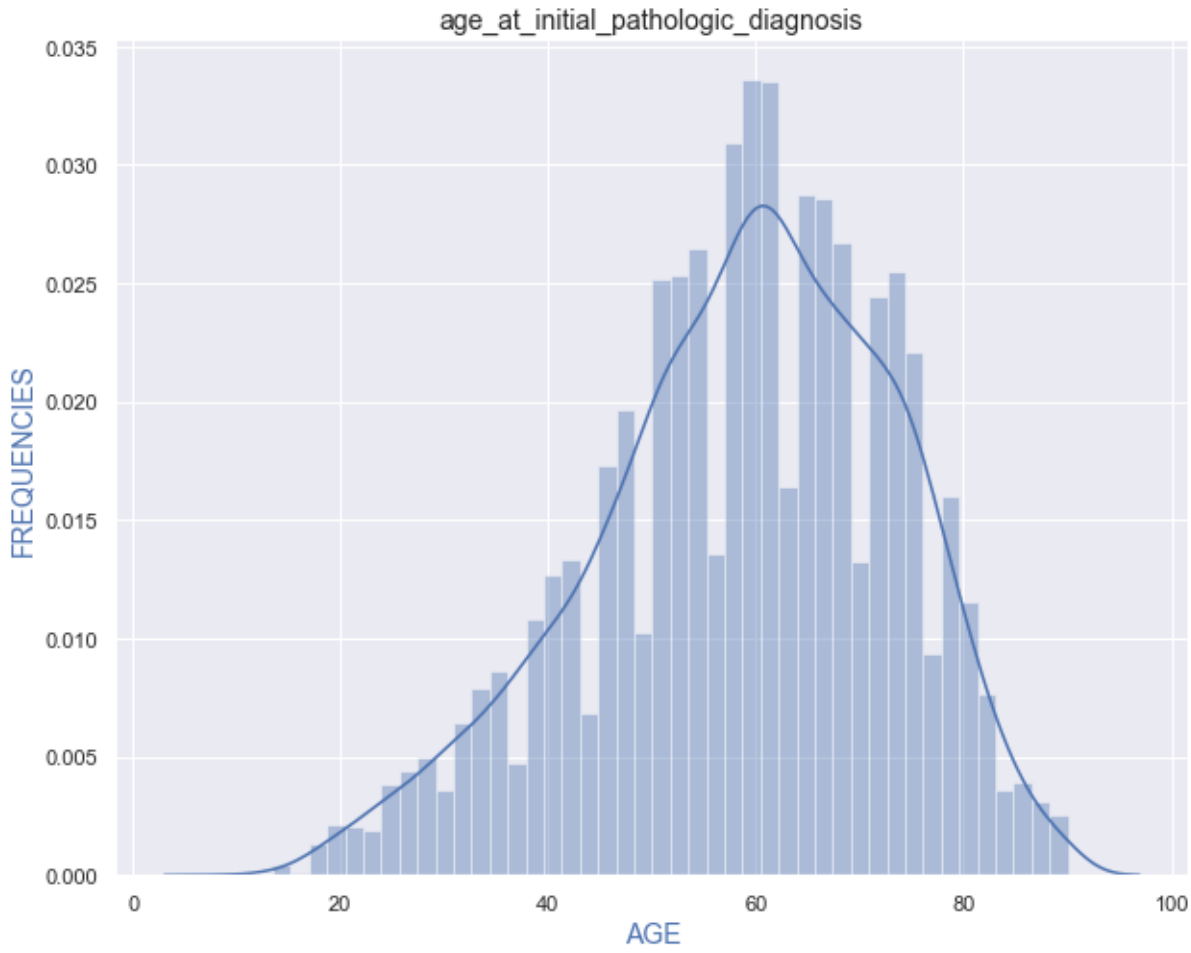
66% of population survived out of 100% and rest 34% didn't survived.



➔ **AGE**

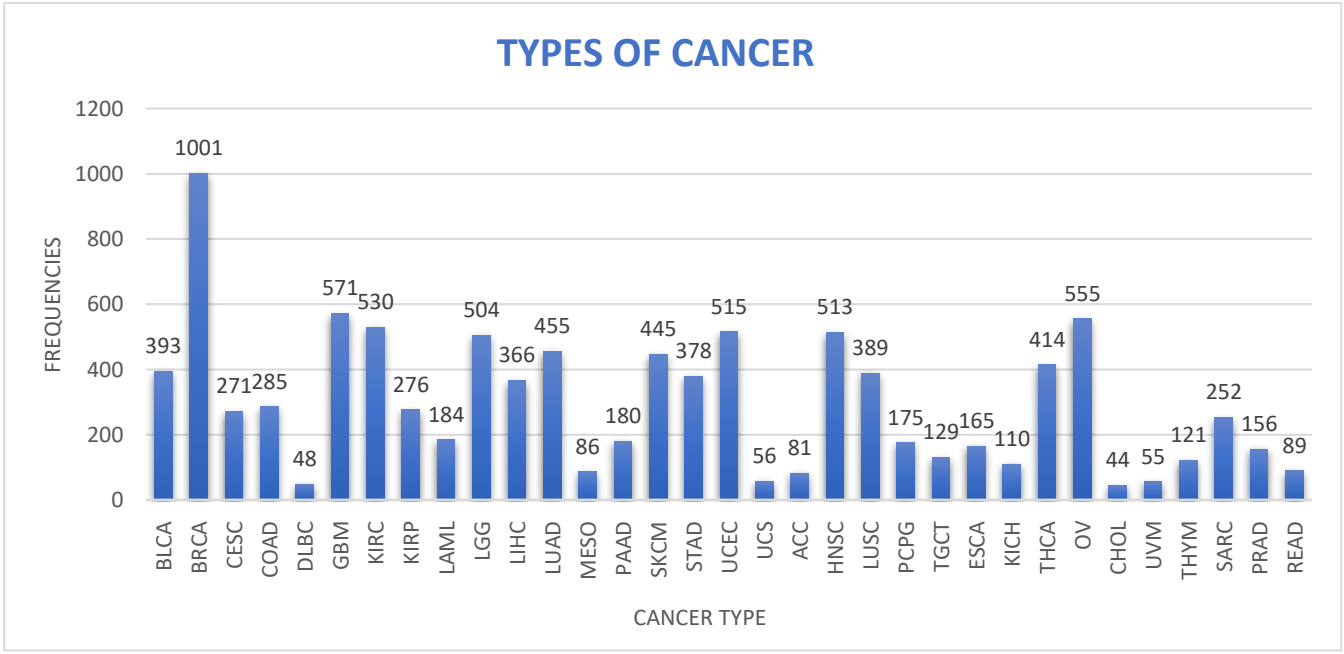
We can see from the following graph that most studied age group is in between 40-80 and it is normally distributed.

Minimum age is – 10 and maximum is - 90

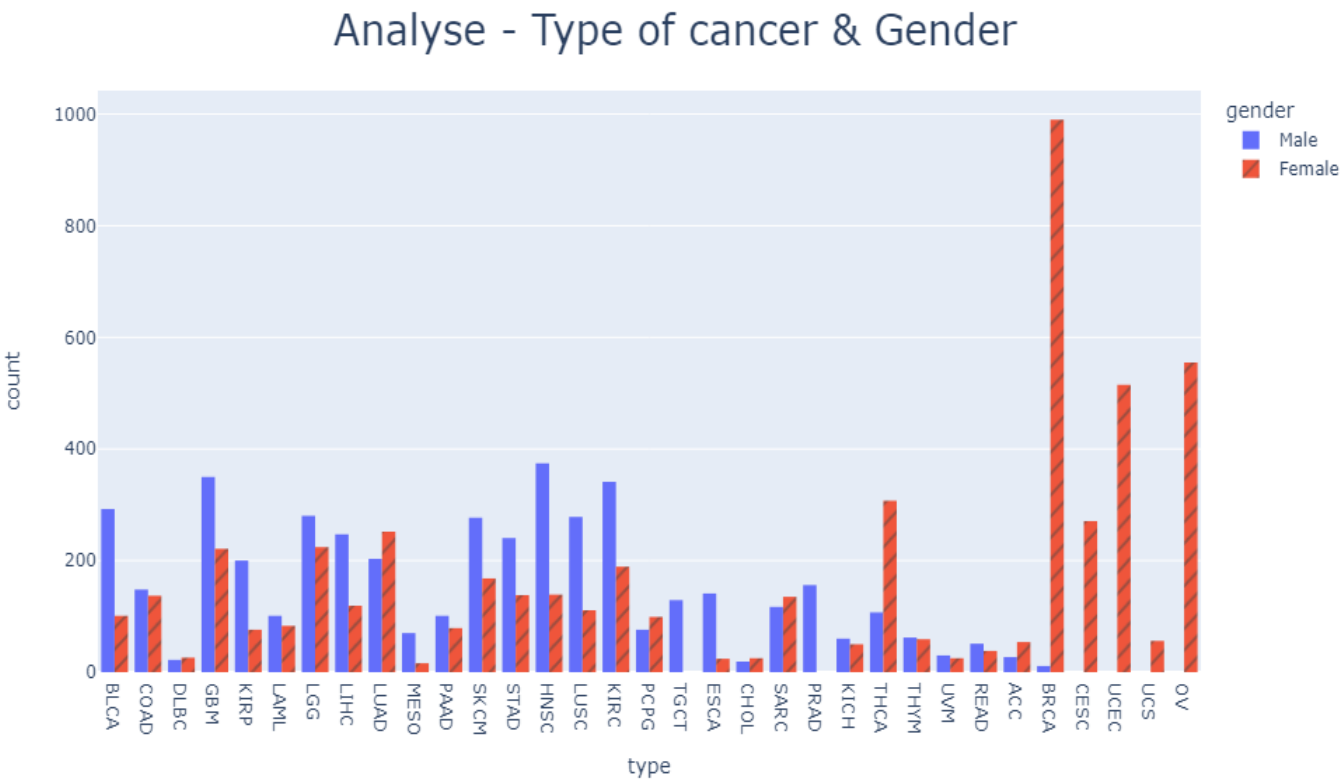


➔ **TYPE OF CANCER**

We have studied 33 types of different cancer.
From graph we can say that most studied cancer out of 33 are BRCA, GBM, KIRC, LGG HNSC, OV. And least studied are DLBC, UCS, CHOL, UVM.



➔ Analyse Gender with respect to different type of cancer



Conclusion -From graph we can say that most prominent cancer in women are BRCA (breast cancer), CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma), UCEC (Uterine corpus endometrial carcinoma), OV (ovarian cancer), THCA (Thyroid cancer).
In males, most prominent cancer are BLCA (Bladder Cancer), HNSC (Head and neck squamous cell carcinoma), LUSC (Lung squamous cell carcinoma), TGCT (Tenosynovial giant cell tumors), PRAD (PROSTATE CANCER).

SURVIVAL ANALYSIS

Now different factors are taken into account to see which group have higher chances of survival.
Different factors are- Gender, Races, type of cancer, age and stages of cancer. Table for the same is described below.

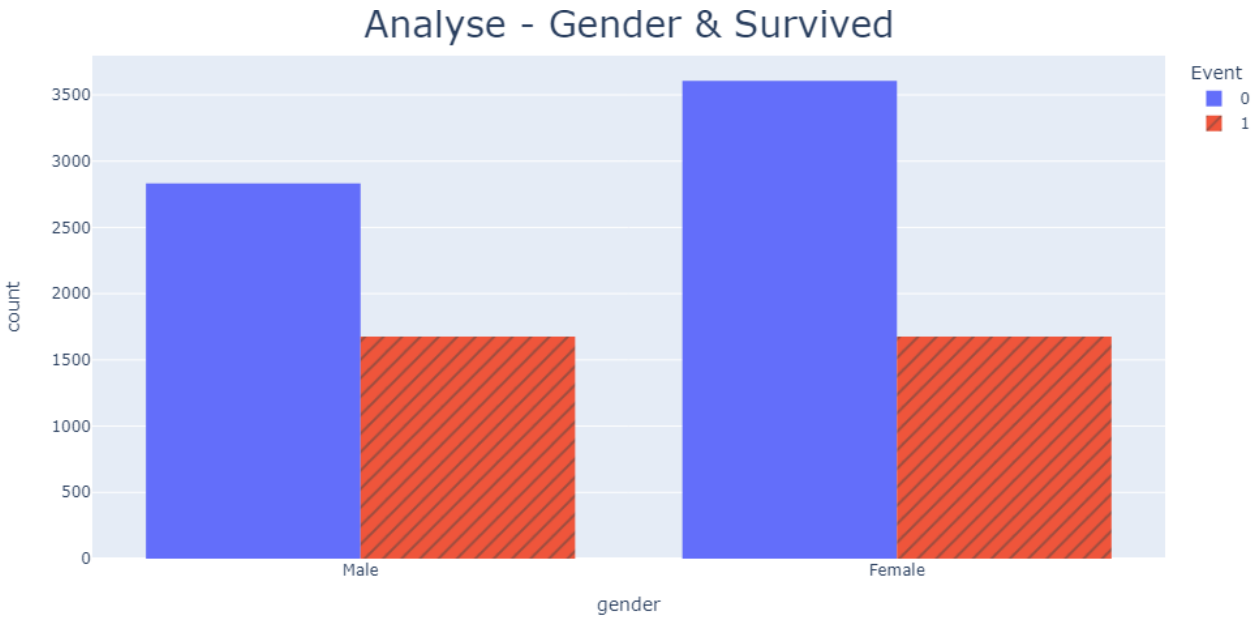
	SURVIVED (0)	NOT-SURVIVED (1)
GENDER		
Male	2833 (0.44)	1677
Female	3606 (0.56)	1676
RACES		
Asian	530	143
White	5224	2927
Black or African Americans	657	271
American Indian	20	7
Native Hawaiian	8	5
TYPE OF CANCERS		
BLCA	220	173
BRCA	857	144
CESC	207	64
COAD	214	71
DLBC	39	8
GBM	100	471
KIRC	335	175
KIRP	233	43
LAML	65	119
LGG	380	124
LIHC	239	127
LUAD	292	163
MESO	13	73
PAAD	84	96
SKCM	233	212
STAD	234	144
UCEC	426	89
UCS	21	35
ACC	49	32
HNSC	296	217
LUSC	216	173
PCPG	169	6
TGCT	125	4
ESCA	105	60
KICH	98	12
THCA	398	16
OV	213	342
CHOL	23	21
UVM	38	17
THYM	112	9
SARC	158	94
PRAD	155	1
READ	72	17
TUMOR STAGE		
Stage I	989	163

Stage IA	287	73
Stage IB	207	114
Stage II	397	161
Stage IIA	553	133
Stage IIB	414	201
Stage IIC	46	20
Stage III	408	214
Stage IIIA	284	174
Stage IIIB	145	84
Stage IIIC	138	82
Stage IV	125	274
Stage IVA	185	132
Stage IVB	9	11
Stage IVC	3	3
Stage X	2	7
IS	43	2
I/II NOS	5	5

TABLE – 2 SUMMARIES OF ANALYSIS

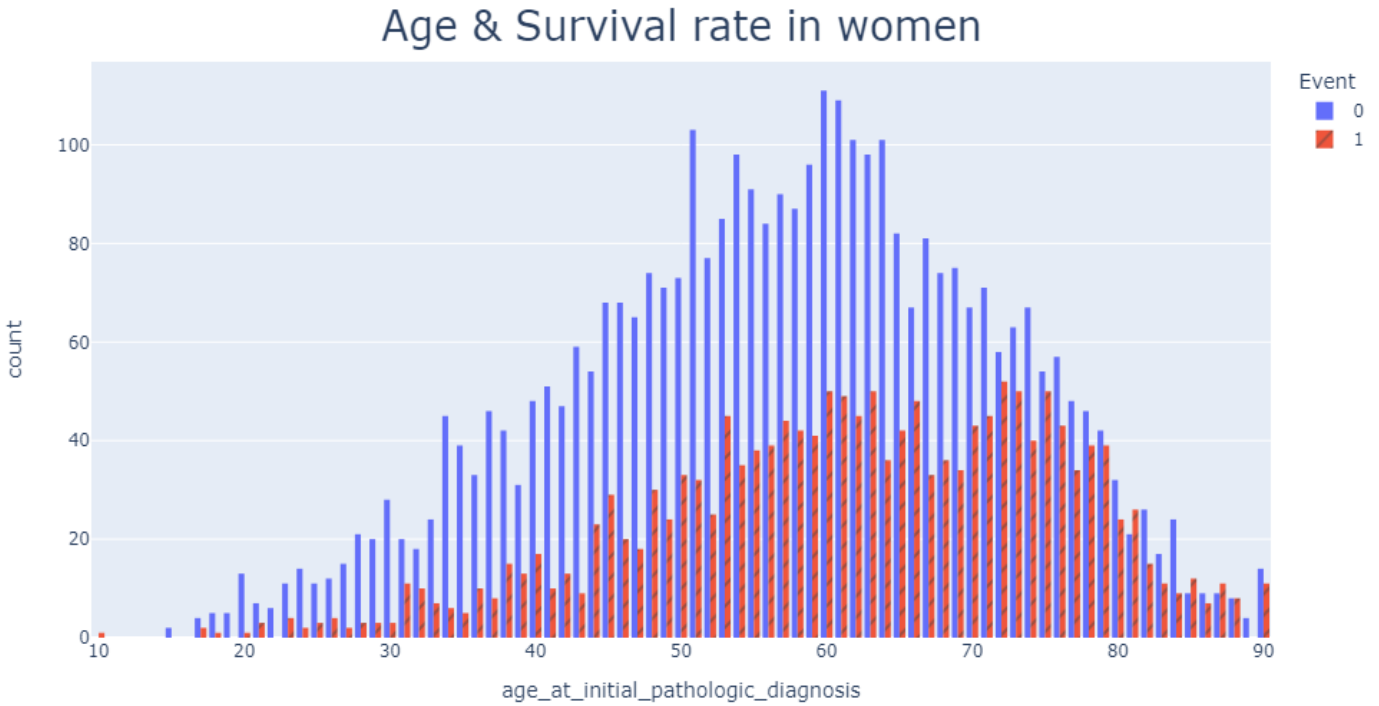
➔ **ANALYSIS OF GENDER WITH RESPECT TO EVENT**

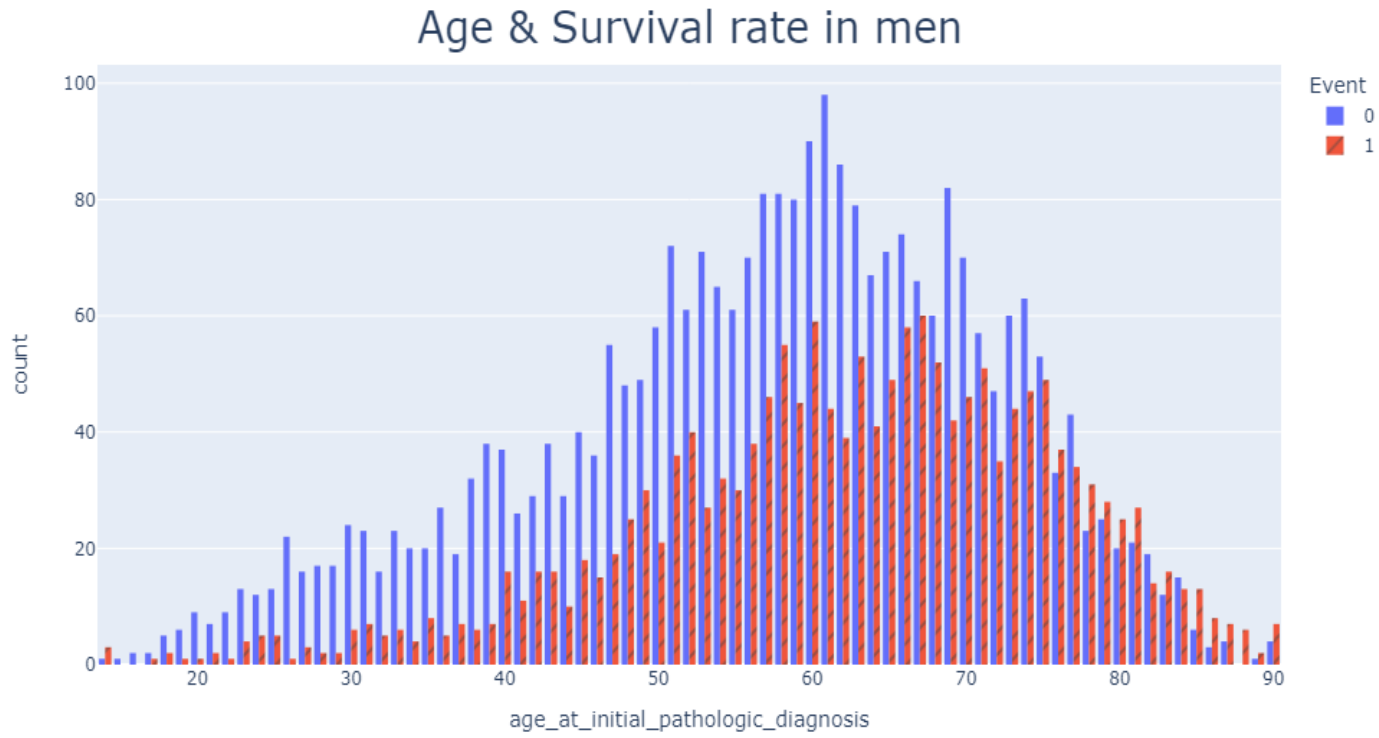
We have found that women displayed better survival rates for the majority of cancer sites



➔ **ANALYSIS OF AGE WITH RESPECT TO EVENT**

In the following graphs, we can see that mortality rate increases with increase in age, there is direct relationship between age and mortality rate.



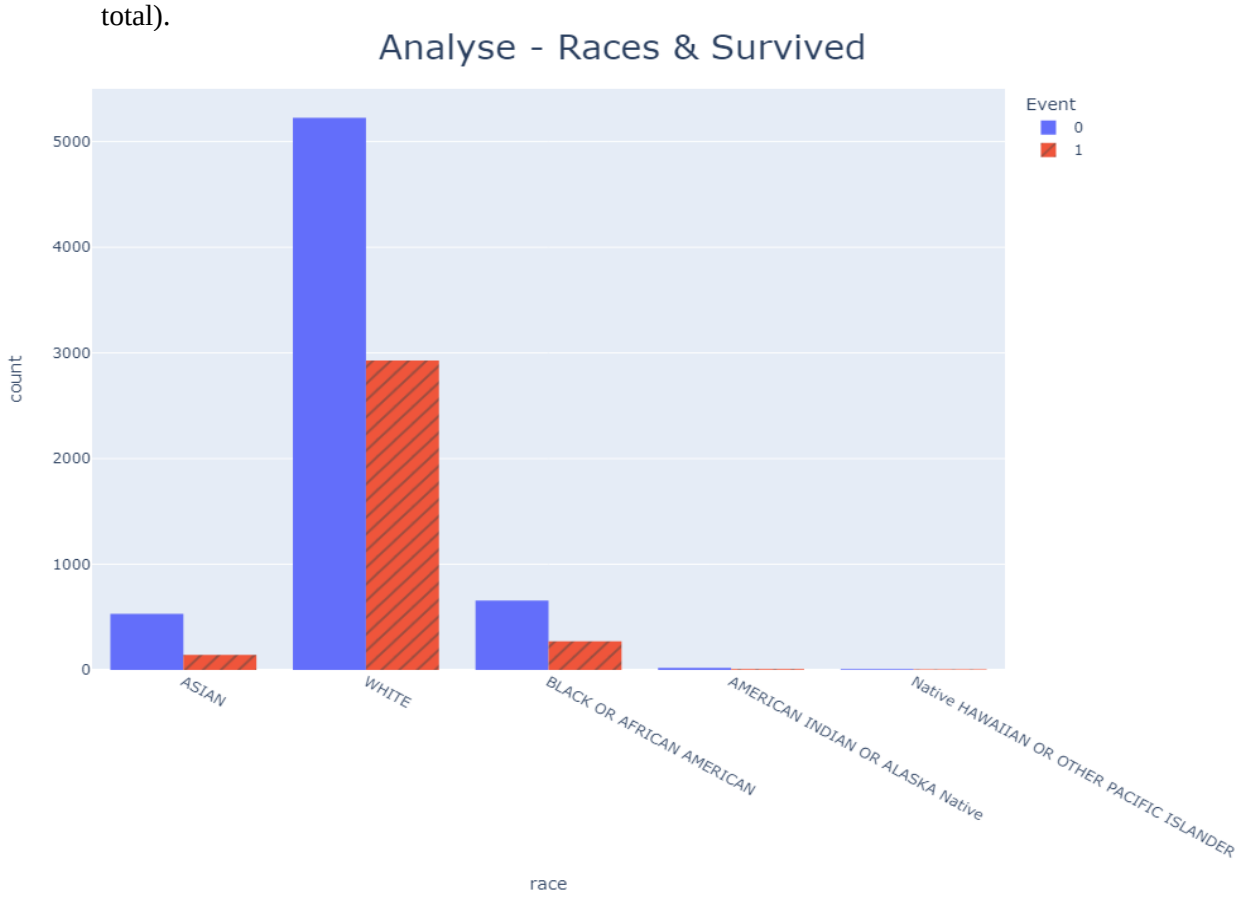


Conclusion- Mortality rate is higher after age 55 (approx.). However, women show low mortality rate as compared to men after age of 55, in women mortality rate drastically increases after 70's. In male, mortality rate increases drastically after 65 (approx.). After age of 80 women have equal survival as well as mortality rate, but men have high mortality rate than survival rate.

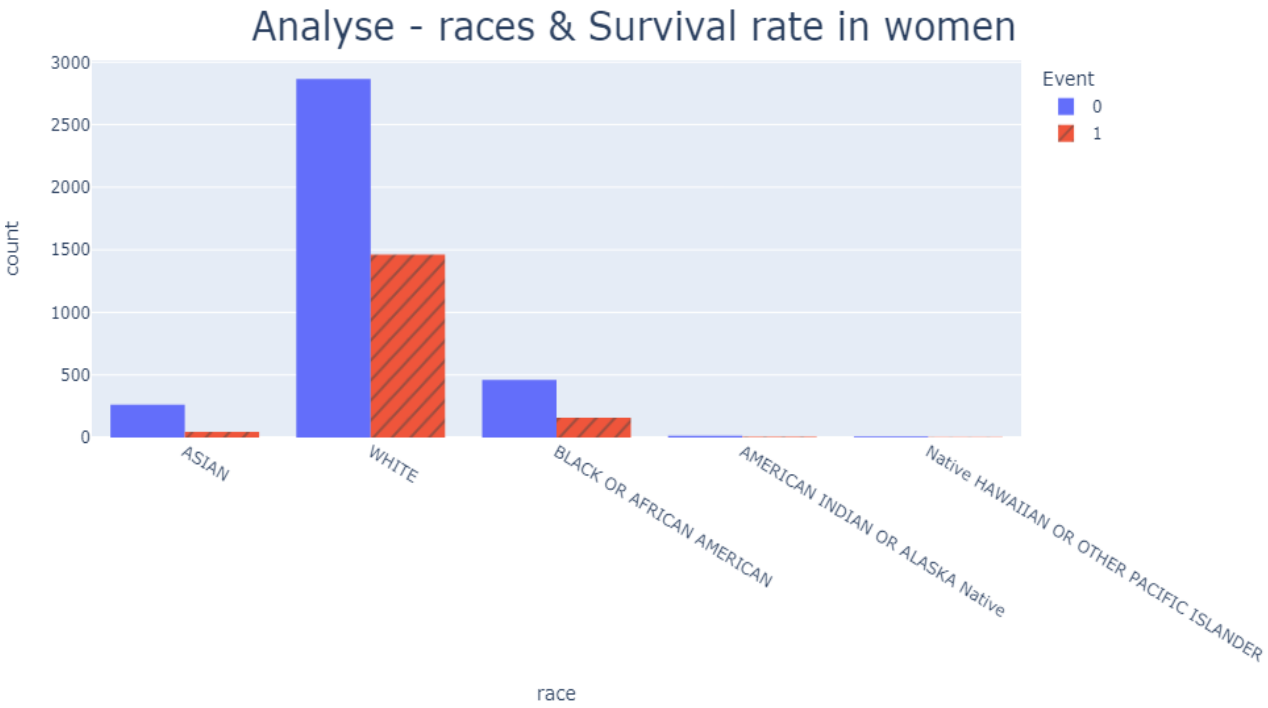
➔ **ANALYSIS OF DIFFERENT RACES WITH RESPECT TO EVENT**

We have studied 5 different races and most of our data is from white race and least is from native hawaiian and american indians.

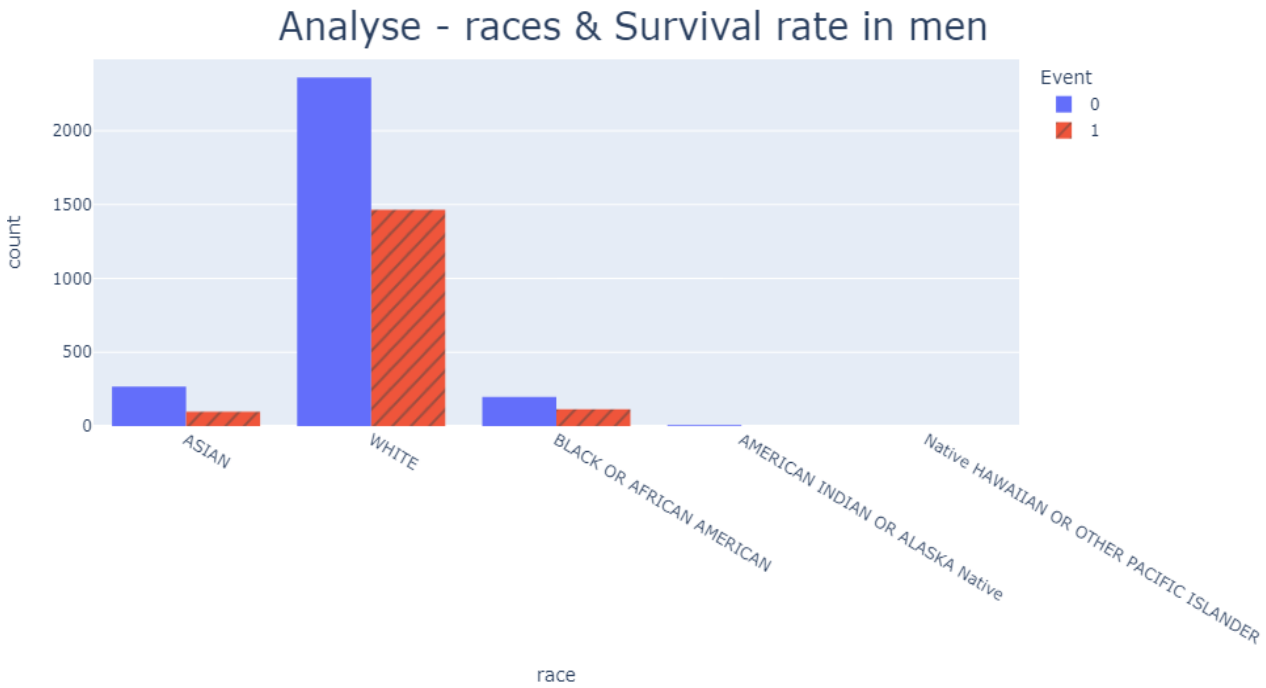
In this case , Almost 79% population survived out of 673 asian population and 64% survived in case of white (8151 total) and 71% survived in case of blacks or africian americans (928 total).



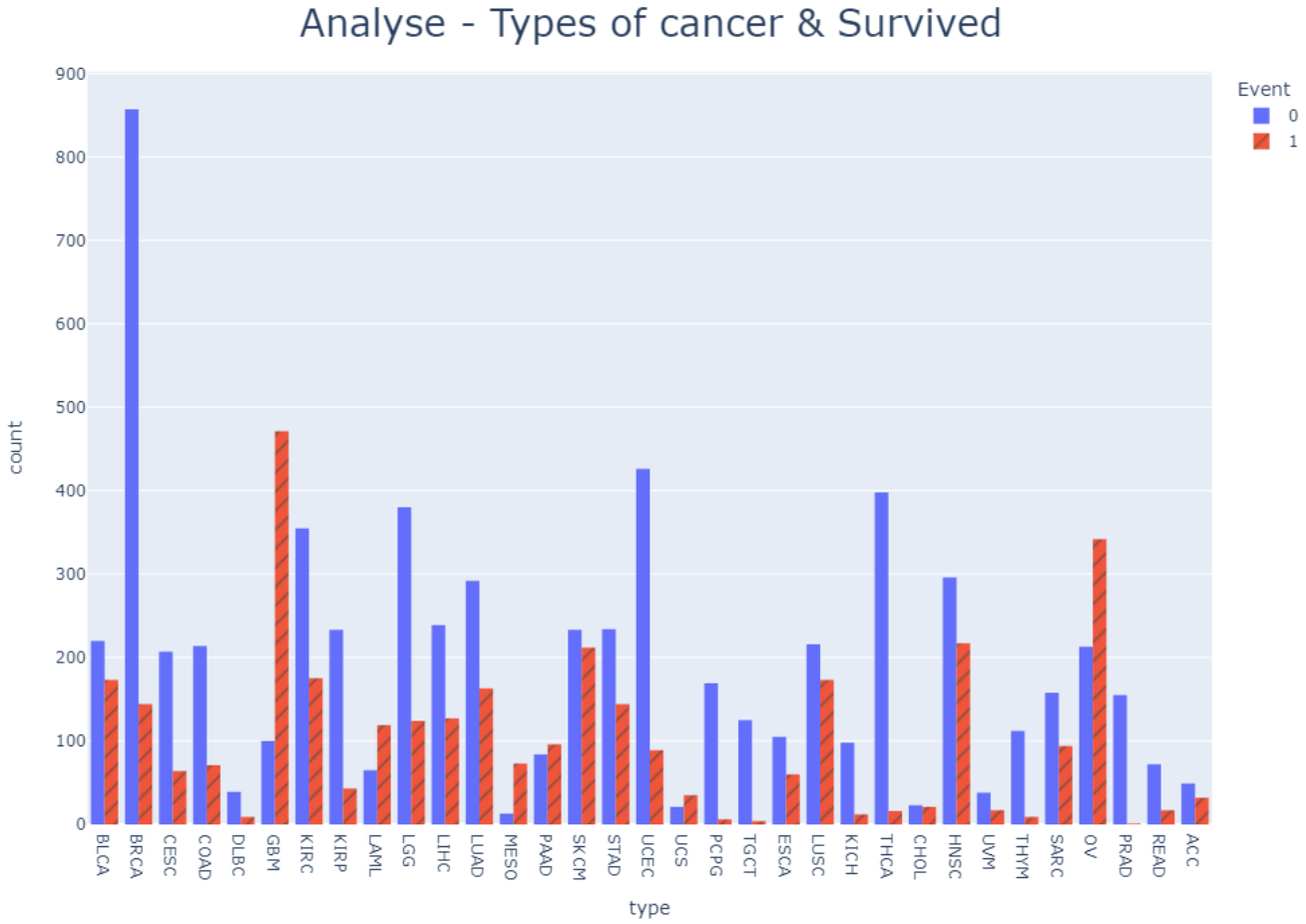
Survival rate of women in different races



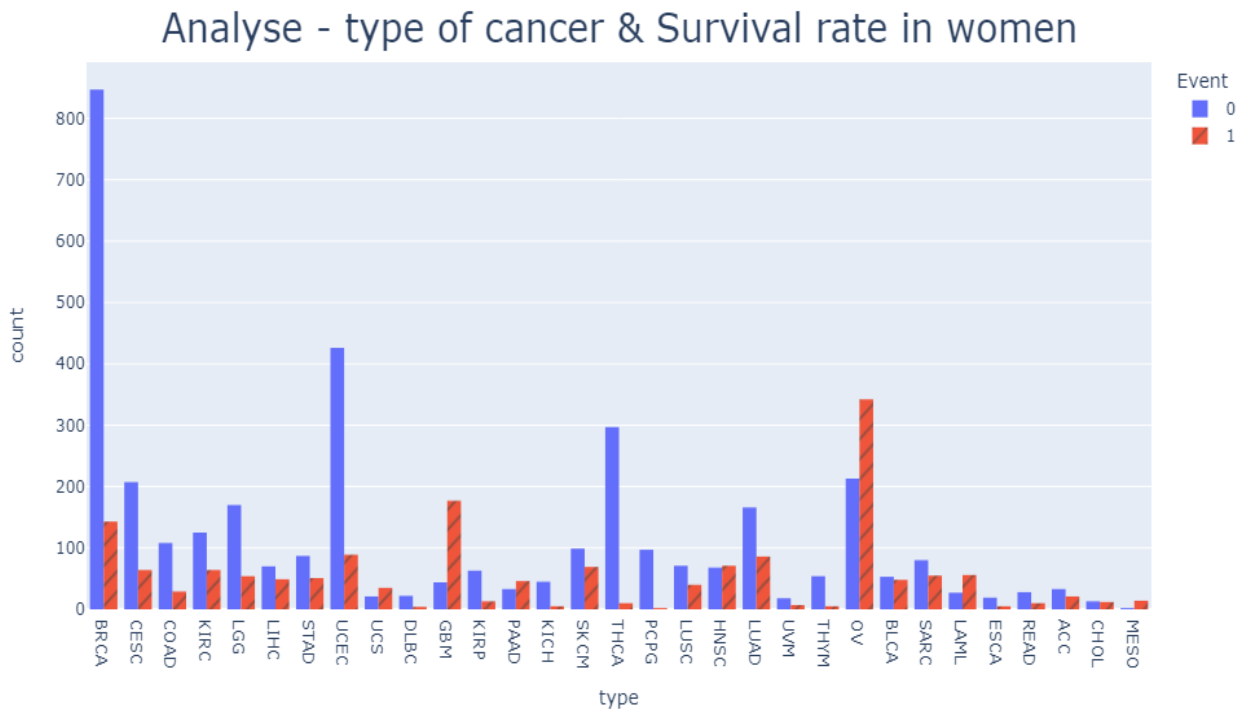
Survival rate of men in different races

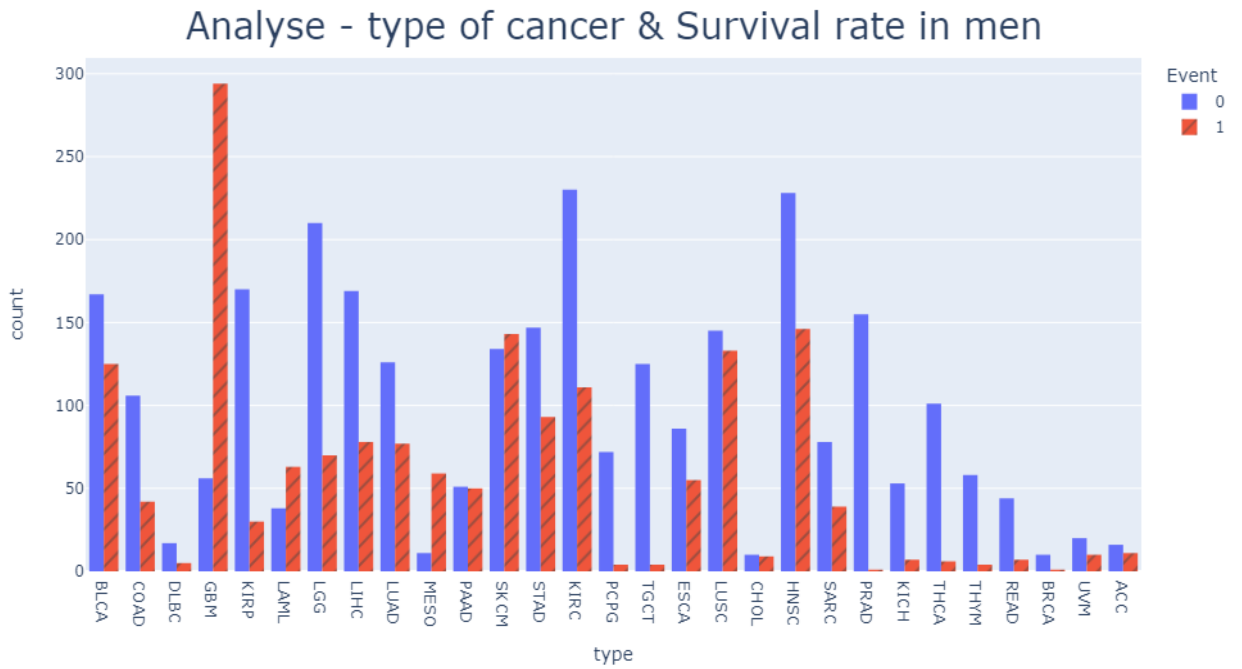


➔ **ANALYSIS OF DIFFERENT TYPES OF CANCER WITH RESPECT TO EVENT**
Survival rate of population with BRCA, UCEC, THCA, TGCT, KIRP, PCPG etc. are more. And mortality rate is high in population with GBM, OV, SKCM, LUSC.



Survival rate of women and men with different type of cancer.





<u>Conclusion-</u> Glioblastoma is an aggressive type of cancer that can occur in the brain or spinal cord and its mortality rate is high in both male and female but comparatively male have higher mortality rate for this cancer than women.

Women have high mortality rate for OV, GBM, LAML cancers. And high survival rate for BRCA, UCEC, THCA,

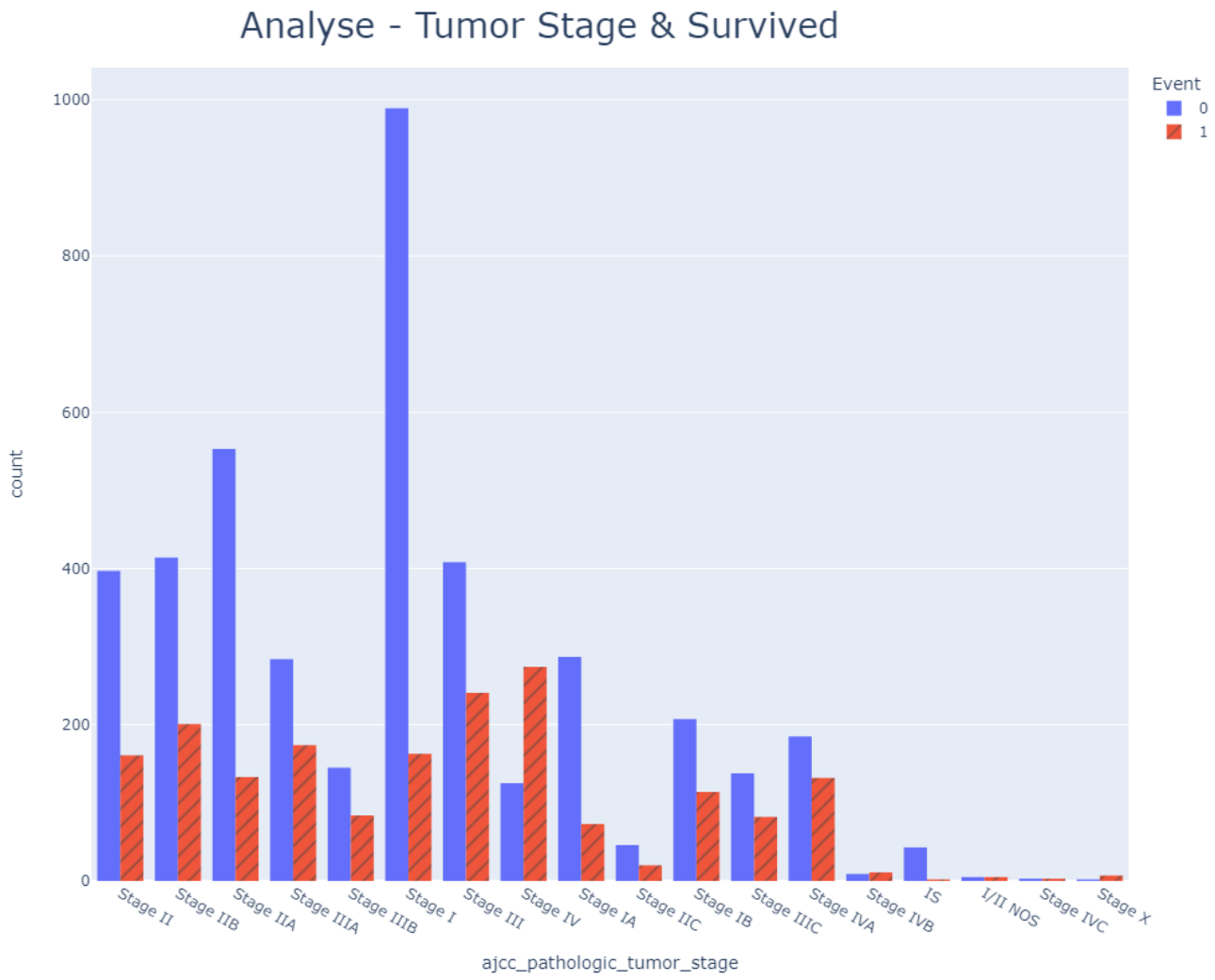
Men have high mortality rate for almost every cancer reported except TGCT, PRAD, THCA, PGPG, THYM.

In short, Women have high survival rates for different cancers as compared to men.

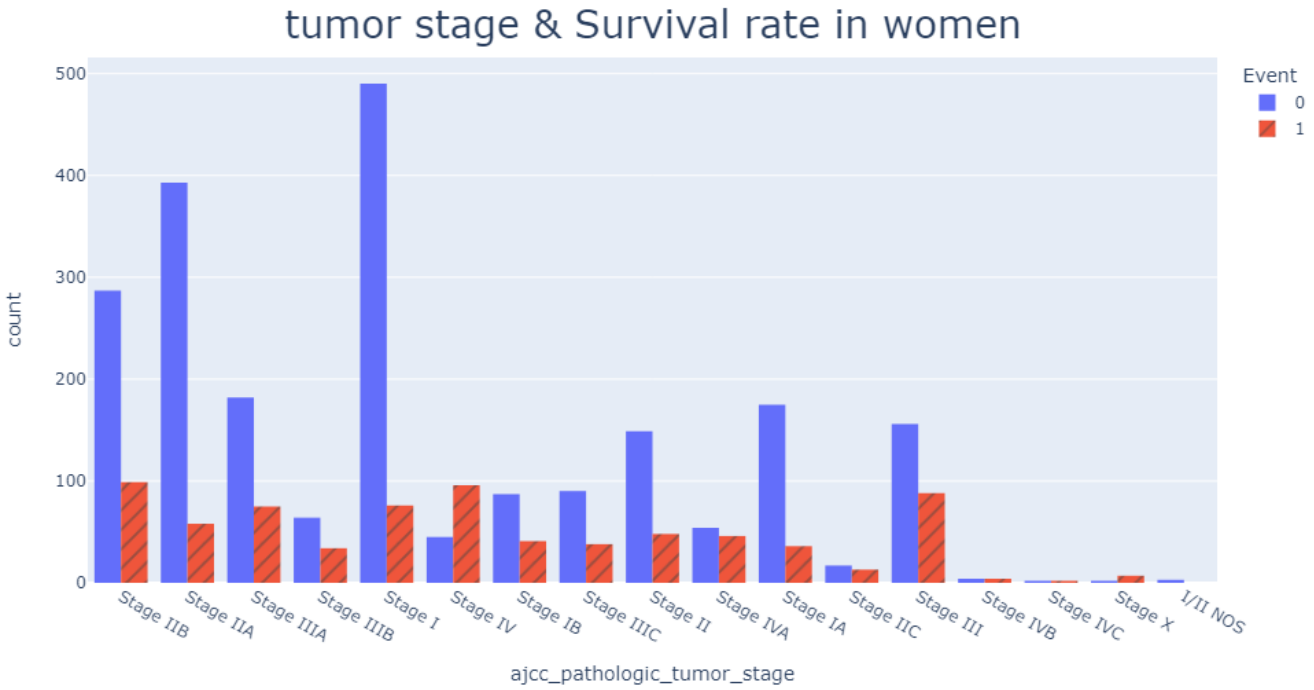
➔ **ANALYSIS OF DIFFERENT TYPES OF TUMOR STAGES WITH RESPECT TO EVENT**

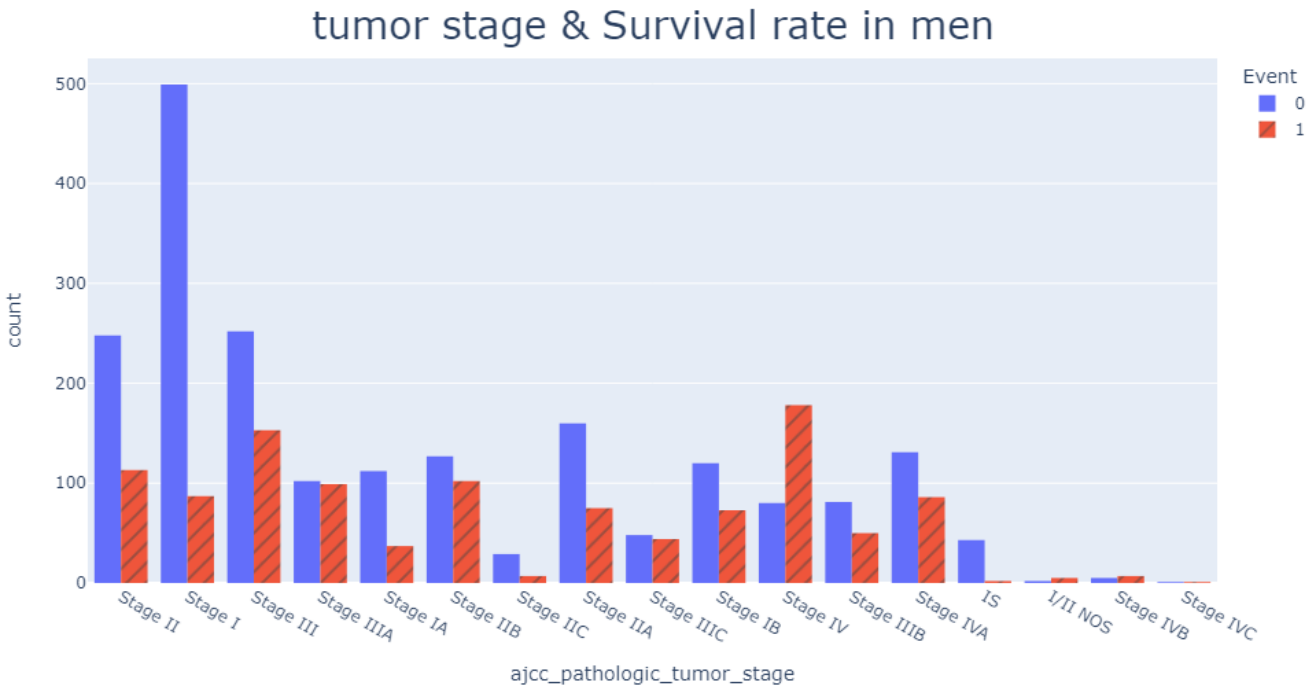
Survival rates are higher when people with cancer detected at stage 1 and stage 2. And mortality rates are higher when cancer detected at stage 4.

- Stage I: Cancer is localized to a small area and hasn't spread to lymph nodes or other tissues.
- Stage II: Cancer has grown, but it hasn't spread.
- Stage III: Cancer has grown larger and has possibly spread to lymph nodes or other tissues.
- Stage IV: Cancer has spread to other organs or areas of your body. This stage is also referred to as metastatic or advanced cancer.



Survival rate of women and men at different stages of tumor.





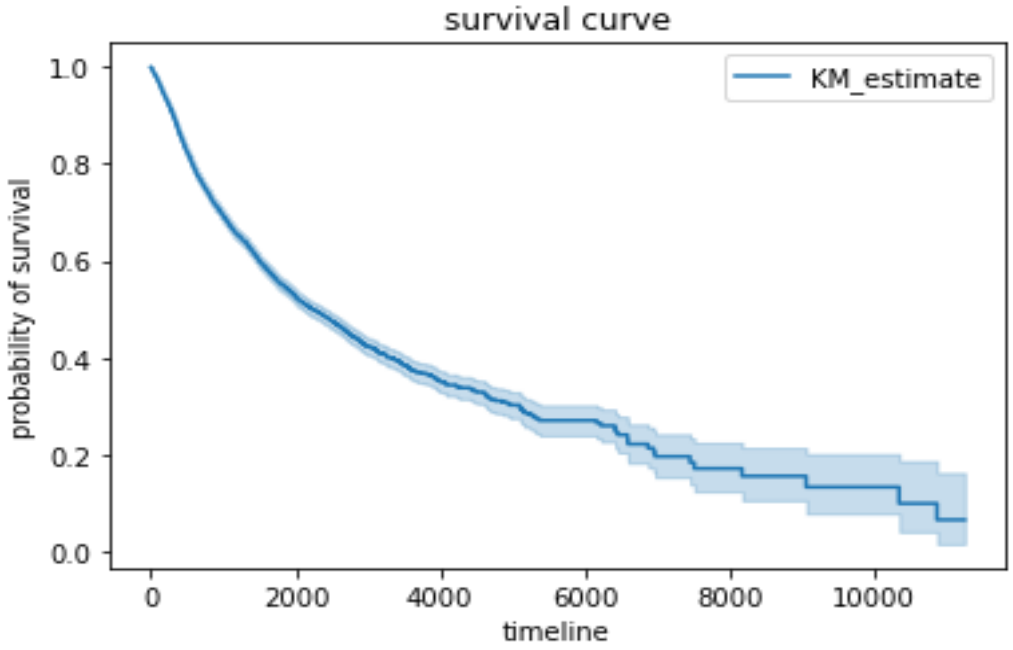
Conclusion- Survival rates of women at different tumor stages are higher than men. However, men also show high survival rate at stage I but overall mortality rate of men is more than women.
Drawback- Our tumor stage data contain various NaN (unknown) value. So, graph was plotted using known value only, but looking at various factor we can interpretate that women survival rate is higher than men. If we removed these unknown values then our dataset will get shorter and many important information will lost and result of analysis will be wrong.

Kaplan-Maier Curve Estimation (Non-Parametric)

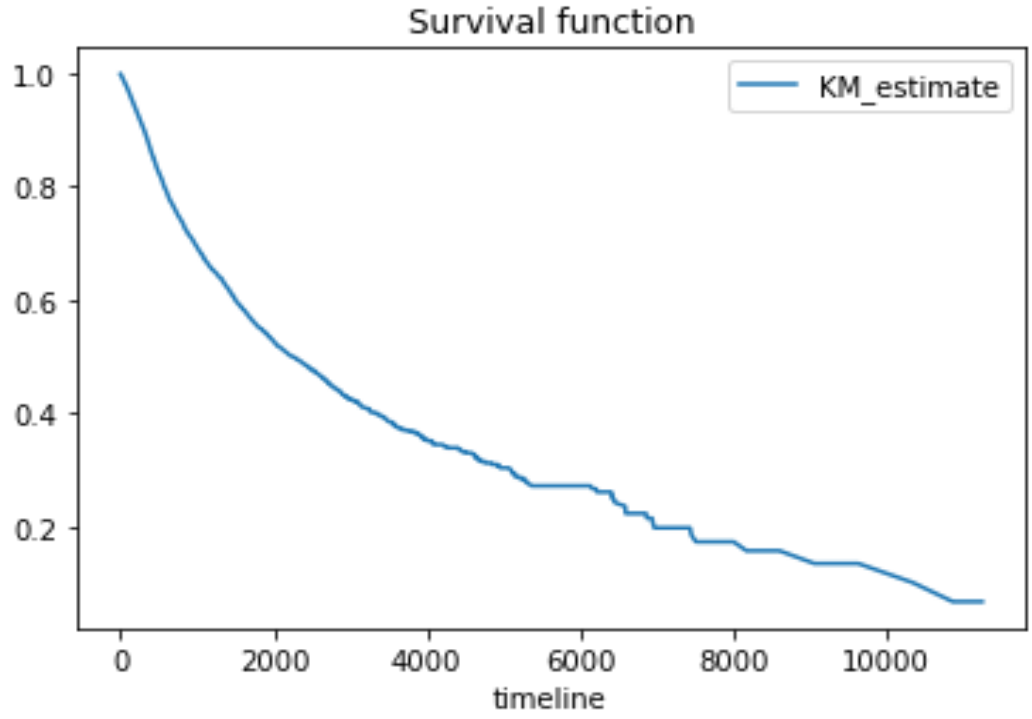
To start with survival analysis, the first step is to plot a survival curve of the overall data. It can be done by generating a Kaplan-Maier curve.
The curve illustrates how the survival probabilities changes over the time. As the time passes, the survival probabilities of cancer patients reduce.

SURVIVAL CURVE

With 95% confidence interval



Without 95% confidence interval

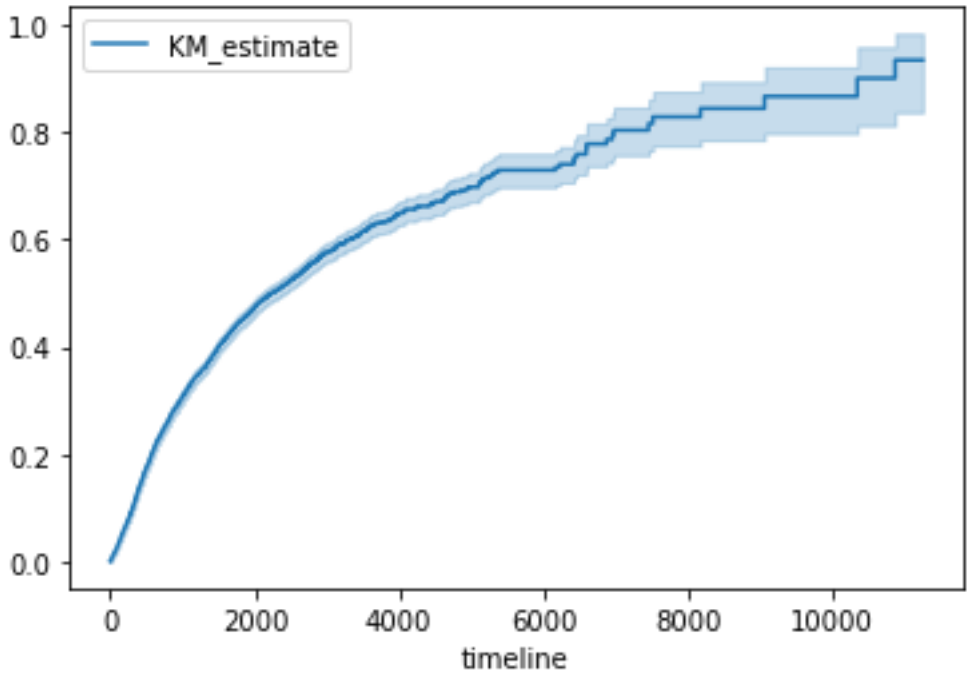


Note- Confidence interval is important it will tell how uncertain we are about each point estimate survival probabilities.
Wide confidence interval means we are less certain.
Narrow confidence interval we are more certain.

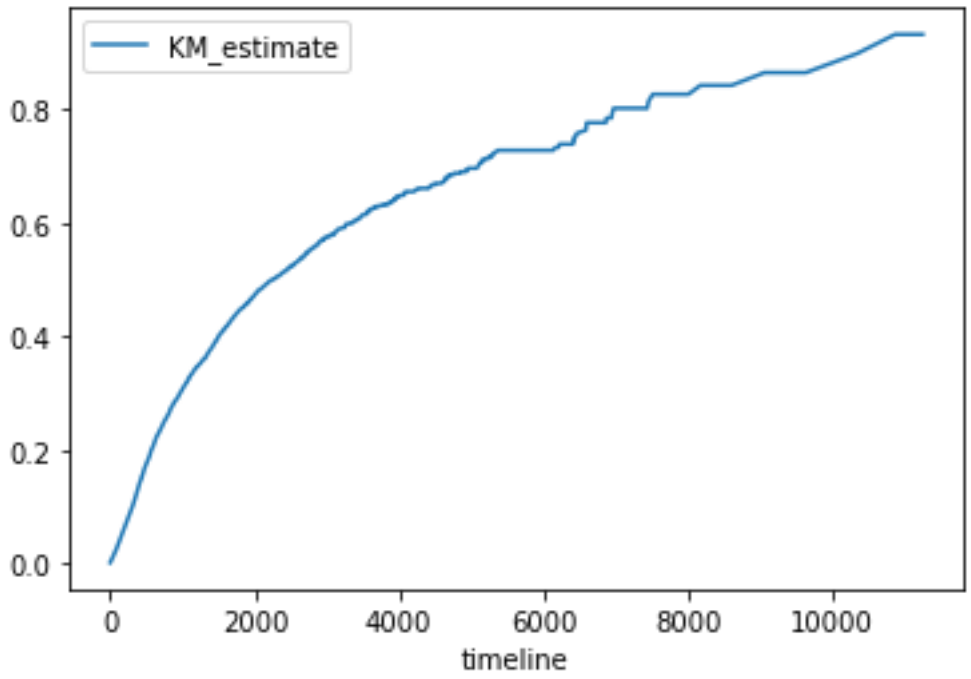
FAILURE CURVE

It is just opposite of survival, i.e., Failure/death probabilities over time.

With 95% confidence interval



Without 95% confidence interval

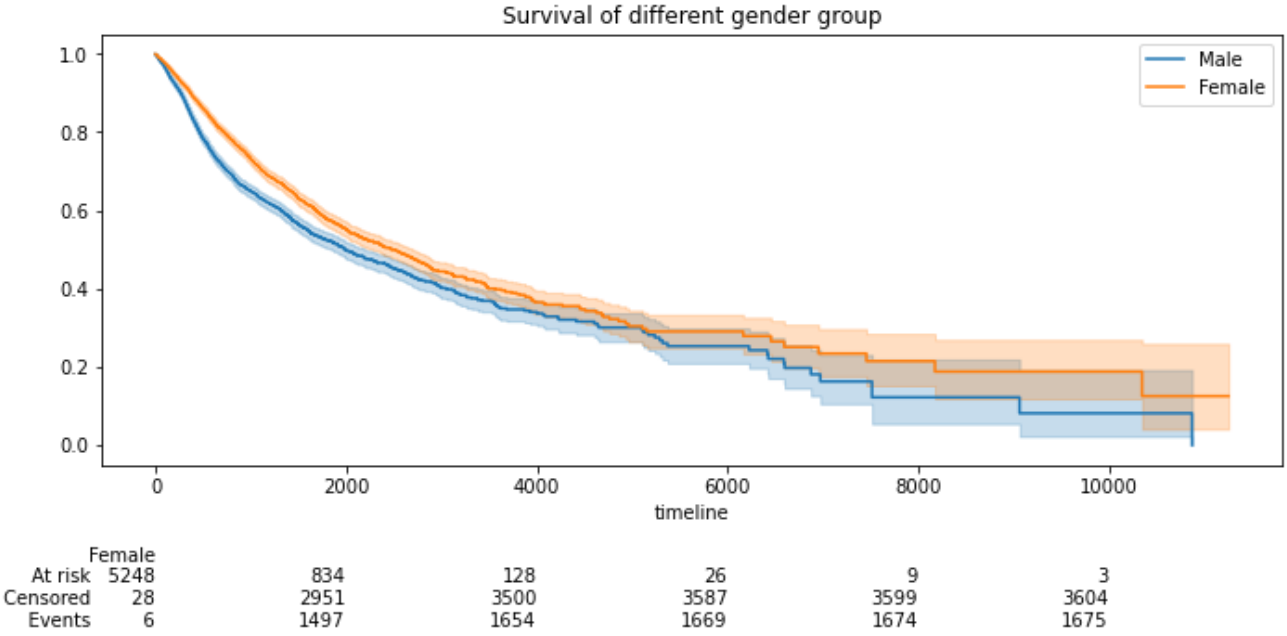


Median - indicates that 50% of the sample live 2240 days and 50% dies within this time.
But most of our data is right censored so we can't estimate median accurately.

```
2240.0
     KM_estimate_lower_0.95  KM_estimate_upper_0.95
0.5                  2097.0                  2417.0
```

Comparing groups' survival distributions

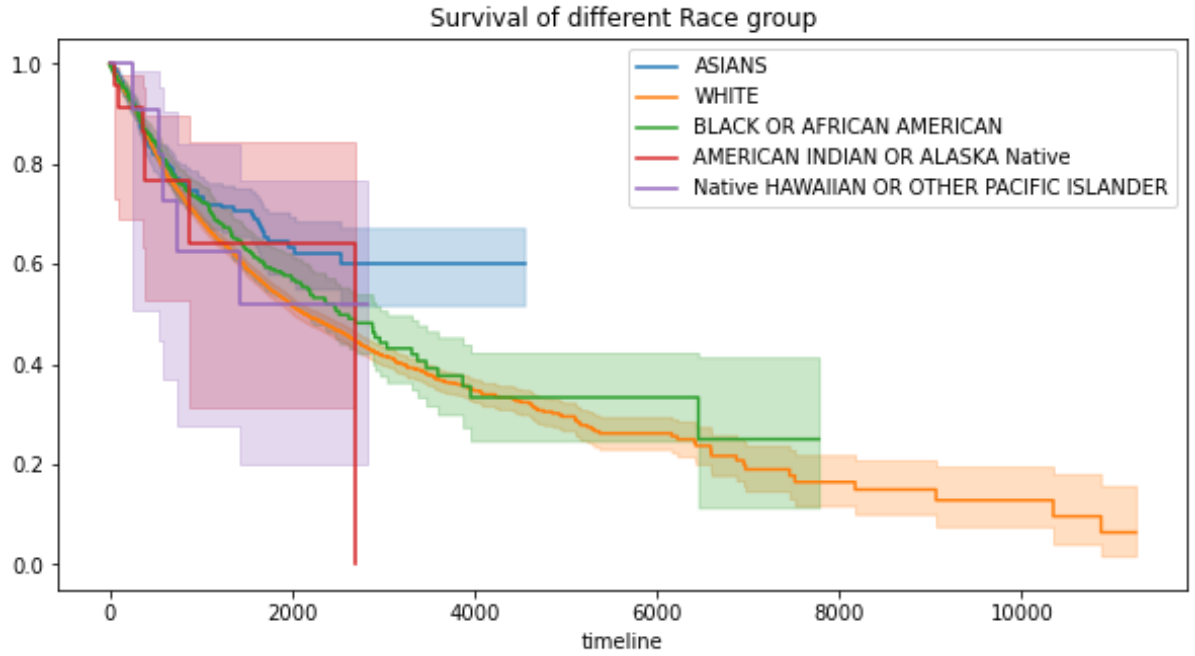
Comparing survival distribution between female and male.



	Female					
At risk	5248	834	128	26	9	3
Censored	28	2951	3500	3587	3599	3604
Events	6	1497	1654	1669	1674	1675

Conclusion- The curve illustrates that the survival probabilities of female patients are overall higher compared to male patients at any instance of time. Here, confidence interval overlaps with each other at some points, it's less likely that there's a real difference between the curves.

Comparing survival distribution between different races.



Conclusion- Here, confidence interval overlaps with each other at some points, so there's not any real difference between the curves. American Indian curve drops to 0 because last value of American Indians in race column would be 1 i.e., not survived.

DISCUSSION

This study among cancer's population of different Races, Age shows that females have higher survival rates than male. In our study females shows survival rate of 0.56 and males shows 0.44.

We have some limitations in our study. First, we did not have information about histologic grade, co-morbidities, or risk factors including smoking. We have only taken factor like age, sex, race, type of cancer into account and predict that mortality rate of Females is low as compared to males.

Possible reason why female have survival rates more than men will be the gene TP53, the most commonly mutated gene in human cancer, with more than half of all human cancers harboring a genetic alteration that interferes with the function of p53 protein. There are some clusters of genes located on the X chromosome that can affect the activity of p53 protein, even if the cancers don't have a mutation in TP53 itself. As we know women contain 2 copies of X-chromosome (XY) and men contain 1 copy of X and 1 copy of Y chromosome (XY).

Therefore, women are less likely to possess mutations in TP53. Second, the presence of p53-regulating genes on the X chromosome means men are particularly vulnerable to defects in these genes. And lastly, extra X chromosome appears to be a barrier in women that prevents the expression of mutant X-linked p53-regulating genes.