[1]:	passenger, and carries information about this passenger such as the passengers's gender, age and whether
[1]:	the passenger survived the Sinking of the Titanic or not. As certain passengers were more likely to survive than others, this dataset can be used for classification tasks. In this project the data is analyzed using python and pandas dataframes.
	Let's have a closer look at the data. # Importing the necessary libraries
	<pre>import pandas as pd import matplotlib.pyplot as plt import seaborn as sns import numpy as np pd.set_option("display.precision", 2)</pre>
[2]:	<pre># Read in data data = pd.read_csv('titanic/train.csv') # Print first five rows</pre>
t[2]:	PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked Braund, Braund, O 1 0 3 Mr. Owen male 22.0 1 0 A/5 7.25 NaN S
	Harris Cumings, Mrs. John Bradley female 38.0 1 0 PC 17599 71.28 C85 C
	(Florence Briggs Th Heikkinen, 3 1 3 Miss. female 26.0 0 0 STON/O2. 3101282 7.92 NaN S
	Futrelle, Mrs. 3 4 1 1 Jacques female 35.0 1 0 113803 53.10 C123 S
	(Lily May Peel) Allen, Mr. 4 5 0 3 William male 35.0 0 0 373450 8.05 NaN S Henry
	The table above displays the first five rows of the dataset. There are 12 columns. Categorical:
	 Survived - 1 if the passenger survived and 0 otherwise. Pclass - the class the passenger travelled on, 1=1st class (most expensive), 2=2nd class and 3=3rd class (least expensive). Embarked - port of embarkation, C = Cherbourg; Q = Queenstown; S = Southampton.
	Sex - the passenger's sex. PassengerId - a unique Id for each passenger. Name - the name of the passenger.
	Ticket - the passenger's ticket number. Cabin - the passenger's cabin. Numerical discrete:
	SibSp - number of siblings or spouses of the passenger that are also on board. Parch - number of parents or children of the passengert that are also on board. Numerical continuous: Age - age of the passenger in years
	Fare - fare paid by the passenger in dollars Problem 1 - Summarizing data in a table
[3]:	To summarize the quantitative data we can use the describe -method of the pandas dataframe. data[['Age', 'SibSp', 'Parch', 'Fare']].describe().transpose()
t[3]:	count mean std min 25% 50% 75% max Age 714.0 29.70 14.53 0.42 20.12 28.00 38.0 80.00 SibSp 891.0 0.52 1.10 0.00 0.00 1.0 8.00 Parch 891.0 0.38 0.81 0.00 0.00 0.0 6.00
	Parch 891.0 0.38 0.81 0.00 0.00 0.00 0.0 6.00 Fare 891.0 32.20 49.69 0.00 7.91 14.45 31.0 512.33 The categorical data is slightly more difficult to summarize using the pandas library.
	For readability the categorical data is summarized in the markdown table below. The code used to find the different values is at the end of the project. Categorical N/891 (%)
	Sex Female 314 (35) Pclass
	1 216 (24) 2 184 (21) 3 491 (55)
	Embarked Cherbourg (C) 168 (19) Queenstown (Q) 77 (9)
	Survived 1 342 (38)
	Problem 2 and 3 - Visualizing data How does passenger class affect survival?
	Anyone who has seen the Titanic movie knows that first class passenger were more likely to survive than third class passengers. Does the data confirm this? Bad plot
[4]:	To answer our question we make a pie chart to visualize the number of survived passengers per class. survival_per_class = data.groupby('Pclass')['Survived'].sum() plt.pie(x=survival_per_class.values, labels=['1st', '2nd', '3rd']);
	plt.title('Number of survived passengers per class'); Number of survived passengers per class 1st
	2nd
	3rd
	Why is this plot bad? Pie charts are generally not the best plot type to use when visualizing data. It is hard to tell the difference in
	size between the different parts of the circle. In this plot it is clear that the 2nd class was the class where the fewest passengers survived. But which is bigger of the 1st class and the 3rd class? By eye it is difficult to compare the sizes.
	More importantly: this plot does not answer our question. We wanted to know if passengers of certain classes were more likely to survive than others. Here the plot shows the total number of survived passengers per class. As the passengers are not evenly distributed on the three classes, this plot does not tell us anything about probability for survival. Looking at this plot you might conclude that class did not
	affect the passenger's survival chance. The next plot will show us that this conclusion is wrong. The pie chart does not reflect the difference in survival rate between the three classes, but rather that there were more passengers travelling by 3rd class than 1st class. Better plot
	The pie chart did not answer our question. We know plot the ratio of passengers that survived per class in a bar plot.
[5]:	<pre>ax = sns.barplot(data=data, x='Pclass', y='Survived', ci=None); plt.ylabel('Percentage survived'); plt.title('Percentage of passengers that survived per class'); Percentage of passengers that survived per class</pre>
	0.6 - 0.5 - 9.2 0.4 -
	0.4 - 0.3 - 0.2 - 0.1 - 0.1 -
	0.0 1 2 3 Pclass Why is this plot better?
	Our plot shows the ratio of passengers that survived per class. Because we calculate the ratio the difference in the number of passengers per class no longer affects our plot. We can now clearly see that 1st class
	passengers were most likely to survive, and that 3rd class passengers were the least likely to survive.
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class.
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots.
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot #Remove outlier data_new = data[data.Fare < 500] plt.scatter(data_new['Fare'], data_new['Survived']);
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot #Remove outlier data_new = data[data.Fare < 500] plt.scatter(data_new['Fare'], data_new['Survived']); plt.xlabel('Fare'); plt.xlabel('Fare'); plt.ylabel('Survived'); plt.title('Scatter plot of survival per paid fare'); Scatter plot of survival per paid fare
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot #Remove outlier data_new = data[data.Fare < 500] plt.scatter(data_new['Fare'), data_new['Survived']); plt.xlabel('Fare'); plt.xlabel('Fare'); plt.xlabel('Scatter plot of survival per paid fare'); Scatter plot of survival per paid fare 10 08
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot #Remove outlier data_new = data[data.Fare < 500] plt.scatter(data_new['Fare'], data_new['Survived']); plt.xlabel('Survived'); plt.xlabel('Survived'); plt.title('Scatter plot of survival per paid fare'); Scatter plot of survival per paid fare 10 Scatter plot of survival per paid fare
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot \$Remove outlier data_new = data[data.Pare < 500] plt.scatter(data_new['Fare'), data_new['Survived']); plt.xlabel('Survived');
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot ###################################
[6]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Remove outlier** ata_new = data[data_paw[*Fare**], data_new[*Survived**]); plt.x.label(*Fare**); plt.x.label(*Fare**); plt.x.label(*Fare**); Scatter plot of survival per paid fare* 10 8 **Gatter plot of survival per paid fare**); Scatter plot of survival per paid fare* 10 8 **Gatter plot of survival per paid fare**); Scatter plot of survival per paid fare**); **Scatter plot of survival
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Remove outlier** data_neve data[data.pare < 500] plt.acatter(data_new['Fare'], data_new['Survived']); plt.tylabel('Fare'); Scatter plot of survival per paid fare **Option of the plot of survival per paid fare'); **Scatter plot of survival per paid fare'); Scatter plot of survival per paid fare observation in the dataset, but it does not tell us anything about how many observations each dot observation in the dataset, but it does not tell us anything about how many observations each dot represents. So as there are many data point in the dataset and there are low-fare passengers who survived and high-fare passengers who old not survive, the scatter plot does not exhibit any relationship between
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot ###################################
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data. I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Removae outlier** data_name_iffered_if_ data_new(!farvived!); plt_claim(!fare'); plt_claim(!fare'); plt_claim(!fare'); plt_claim(!fare'); plt_claim(!fare'); plt_claim(!fare'); plt_claim(!fare'); Scatter plot of survival per paid fare **Open passengers who survival per paid fare'); **Scatter plot of survival per paid fare **Open passengers who survival per paid fare to be more likely to survive but there are also low-fare passengers who survive. The scatter plot displays a dot for each distinct observation in the dataset, but it does not tell us anything about how many observations each dot represents. So as there are many data point in the data set and there are low-fare passengers who survived and high-fare passengers who survive, the scatter plot does not exhibit any relationship between paid fare and probability of survival. Better plot - a box plot **san boxplot(data=data_new, x='Survivad', y='Sara')) plt_cliie('Face paid by passengers per class) **Face plot + a box plot
	Additionally it is easier to compare the helights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the faire affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class also paid a higher faire than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in faire. This is an extreme outlier.\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Removal outlier** data_pew = data(data_trace < 5001) plit. valued('Pare'); plit. valued('Pare'); plit. valued('Pare'); plit. valued('Survivad'); plit. valued('Survivad'); plit. valued('Survivad'); plit. valued('Survivad'); plit. valued('Survivad') and paid fare'); **Scatter plot of survival per paid fare') **Scatter plot of survival per paid fare') **Scatter plot of survival per paid fare'); **Scatter plot of surv
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-asis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3nd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3nd class. People in 1st class also paid a higher fare than the people in 2nd and 3nd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier. As this outlier stretches the scale of the plots and makes it hander to see the distribution of the rest of the data. I have removed this outlier when creating the following plots. Bad Plot - scatter plot ###################################
	Additionally it is easier to compare the helights of the bars than it is to compare the size of pie parts. The numbers on the y-asis make it possible to read the actual sunvival probability of each class. Just by locking at this plot we can tell that the sunvival probability was about 47% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3nd class. And that a passenger on 1st class was more than twice as likely to survive than the people in 3nd class. People in 1st class also paid a higher fare than the people in 2nd and 3nd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier. As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Remove outlier** data_num = data** data** data.** Agra < 501** ppt. value** [value**] part = value** [value**] part = part = value** [value**] part
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability and such drifts for passengers what yo looking at this plot we can tell that the survival probability as about 47% for passengers on the 3rd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. Poople in 1st class also paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier. As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Removal** outlier** data of a stretches the closwing plots. Bad Plot - scatter plot **Removal** outlier** data of a stretches the closwing plots. Bad Plot - scatter plot **Why is the scatter plot bad?** It is hard to see pattern in this plot. We expect passengers who paid a high fare to be more likely to survive but there are also low-fare passengers who survive. The scatter plot displays a dot for each distinct observation in the dataset, but it does not tell us anything about how many obsenations each dot represents. So as there are many data point in the dataset as each other ace low-fare passengers who survived and high-fare passengers who did not survive, the scatter plot does not exhibit any relationship between paid fare and probability of survival. Better plot - a box plot **The scatter plot showed us that there are both low-fare and high-fare passengers who survived, and both
	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on the y-axis make it possible to read the actual survival probability of each class. Just by looking at this plot we can tell that the survival probability was boat of 78% for passengers travelling on 2nd class, and that a passenger on 1st class was more than twice as likely to survive as a passenger on the 3rd class. How does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. People in 1st class along paid a higher fare than the people in 2nd and 3rd class. So what is the relationship between fore and probability for survival? Removal of outlier There is a passenger who paid \$1512 in fare. This is an extreme outlier. As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier when creating the following plots. Bad Plot - scatter plot **Pressive usufface** **Pressi
[7]:	Additionally it is easier to compare the heights of the bars than it is to compare the survoil pie parts. The numbers on the y-soks make it possible to read the actual survival probability of each class, but by looking at this plot we can be that the survival probability was about 47% for passengers threeling on 3rd dass, and that a passenger on 1st class was more than twice as itsely to survive as a passenger on the 3rd class. About does the fare affect survival rate? From the previous plot we know that people in 1st class were more likely to survive than the people in 3rd class. Propie in 1st class, sho paid a higher fore than the people in 7nd and 3rd class. So what is the relationship between fare and probability for survival? Removal of outflier R
[7]:	Additionally it is easier to compare the heights of the bast than it is to compare the size of pie parts. The numbers on the y-usis make it prosible to read the actual survival probability of each class, Just by plotsing at this plot we can cell that the survival probability was obtained at this plot we can cell that the survival probability was obtained at the provided probability and the passenger on the 2rd class, and drast passenger on 1st class was more than twice as likely to survive as a passenger on the 2rd class, and that a passenger on 1st class, also paid a liagher fare than the people in 2rd class. How does the fare affect survival rate? From the previous plot we know that passenger who pied \$512 in fare. This is an extreme outlier\ As this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data, I have removed this outlier where creating the following plots. Bad Plot - scatter plot ### Section outlier of the plots and probability for survival: ### Section outlier outlie
[7]: [8]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The mumbers on the yours make it possible to read the actual survival public place of each days, but by plotsing at this plot we can that the survivals probability was about 47% for passenger trivelling on 2nd diss, and that a passenger on 1st dass was more than twice as likely to survive as a passenger on the 3nd dass. How does the fare affect survival rate? From the provious plot we know that people in 1st dass was more than two people in 2nd and 3nd dists. People in 1st dass also paid a higher fare than the people in 2nd and 3nd dists. So what is the relationship between five and probability for survival? Removal of outlier There is a passenger who paid \$517 in fare. This is an extreme outler \(\text{A}\) this outlier stretches the scale of the plots and makes it harder to see the distribution of the rest of the data. I have removed this outlier when creating the following plots. Bad Plot - scatter plot ### Scatter plot as survival per paid fare #### Use the scatter plot distance of the plots are scatter plot distance in the distance of the plots are scatter plot as survival. The scatter plot distance is the scatter plot distance of the plots are scatter plot distance of the plots are scatter plot distance in the distance but if does not eshability to survival the scatter plot distance in the distance but if does not tell us mything about how many observations each dat represents. So a three are many data opin in the data set and three are low for passengers who survival and high-fare passengers who survival and high fare and high the real many data and the plots of the survivance of the plots and high fare passengers who survival and high fare passengers who survival and high fare of the survivance high plots of the survivance horizons and there are outlies from passengers who there ar
[7]: [8]: [9]: [10]:	Additionally it is easier to compare the heights of the bars than it is to compare the size of pie parts. The numbers on they ask make it possible to read the actual survival probability of each data. Just by locking at this plot we can the the the survival probability was broad left for prossingers broading on Part desix, and that a passenger on its class was manch than toke as likely to survive as a passenger on the fact data. How does the fare affect survival rate? Time the prossup that we know that people in but class were more flarly in survive than the people in that class. No what is the relationship between free and probability for survival? Removal of outlier There is a passenger who paid \$512 in fare. This is an extreme outlier. As this outlier stretches the scale of the peas and mains it handor in see the destination of the rest of the data. These removed this multier of the peas and mains it handor in see the destination of the rest of the data. These removed this multier of the peas and mains it handor in see the destination of the rest of the data. These removed this multier of the peas and mains it handor in see the destination of the rest of the data. These removed this multier can be peased to the pease and mains it handor in see the destination of the rest of the data. These removed this multier of the pease and mains it handor in see the destination of the rest of the data. These removed this multier of the pease and mains it handor in see that destination of the rest of the data. These removed the multiple of the peased in the data and the peased in the data as and the peased in the data as and the peased in the peased in the section of the data and the peased in the data as and the peased of the peased in the data as and the peased of the peased in the peased in the data as and the peased of the peased in the peased in the data as and the peased in the peased in the data as and the peased of the peased in the peased i
[7]: [8]: [9]: [10]:	Additionally it is easier to compare the helights of the bast shan it is to compare the size of pie parts. The morters on they asso make it possible to read the actual survivel probability of each class. Just by booking at this place were the first the survivel probability was short 25% or presented the class. Just by booking at the place of the first the purities probability was short 25% or presented the class. All this apparency on 15t class was more than toxe as thely to survive as a passenger on the 3rd class. The first and the present of the class also paid a higher first than the people in 2nd and 3rd class. So what is the relationship between face and probability for survival? Removal of outlier There is a passenger who paid 312 in face. This is an extreme outlier, As Just outlier suitable the scale of the place and makes the head or see set be distribution of the rest of the class, These removes this outlier when creating the following place. Bad Plot - scatter plot **Section of States and Compare the scale of the class of the passenger who paid 312 in face. This is an extreme outlier, As Just outlier suitable the scale of the place and makes the head or see set be distribution of the rest of the class, These removes this outlier when creating the following place. **Bad Plot - scatter plot **Section of States and Compare the scale of the class of
[7]: [8]: [9]: [10]:	Additionally it is easier to compare the heights of the basis than is tor compare the view of pit parts. The numbers on they axis make it provides to provide the trade private processing view and a for provides or they have the provides of the provides provides of the p
[7]: [8]: [9]: [10]: [11]:	Additionally it is existed to compare the heights of the bass shart is to compare the size of perparts. The numbers on they also make it possible to each distance plants in the possible to seed the acut is underly probability of each class. Just by looking at the plants of the that the survival probability was about 47% of passengers treating on and class, and that a passenger on 1st class was more than note as filled by a survive as a passenger on 1st class was more than note as filled by a survive and a passenger on 1st class was more filled by a survive and a passenger on the bid class. How the probability for survival is the resistanch persons from any probability for survival is the resistanch persons from any probability for survival. **Removal of outlier** There is a passenger who pad \$12 in ferr. This is an extreme outlier. As this outlier sireshes the scale of the plants and make it harder to see the distribution of the rest of the class. I have removed this outlier sireshes the scale of the plants and make it harder to see the distribution of the rest of the class. I have removed this outlier sireshes also solve the plants and make it harder to see the distribution of the rest of the class. I have removed this outlier sireshes also solve the plants and make it harders that the distribution of the rest of the class. I have removed this outlier sireshes a solvent that the see that the plants are probable to see that the class of the class o
[9]: t[9]: [10]:	Additionally is dealer to compare the heights of the bars than it is to compare the size of die parts. The numbers on they aim make it protects to read the actual invoval probability of each date, but by locking which side on core and than the visuroly perceits invoval probability of each date, but by locking and than a promiting me for diese was more than before sidely to survive as a prevenger on the fact date. How does the fare affect survival rate? From the produce plot we know that profile in fact does ever more fliely to survive than the people in Sed date. Beach in 1st date shap paid a higher for them the propole in 2nd and 2nd date. See what is necessitably became in ear probability for survival for the profile in 2nd and 2nd date. See what is necessitably became in ear probability for survival for the profile in 2nd and 2nd date. See what is necessarily became in ear probability for survival for the sea of the date. There is a survival for the sea of the date of the date of the sea of the sea of the date of the date of the control of the control of the date of
[8]: [7]: [10]: [10]: [12]: [13]:	Additionally in coales to compare the neights of the base trans it is compare the size of pip pars. The numbers or they was make it processed to reach the actual seriously probability of each class, as it by to be a compared to the coales and the seriously processes are presented to the coales and the seriously between the and people in the days were more likely to service that the people in the days when the seriously between the and people in the days were reached to the seriously between the and people in the days were reached to the seriously between the and people in the days were the seriously between the and people in the days were the people in the days were reached to the seriously between the and people in the days were the people of the seriously between the and people in the seriously and the seriou
[7]: [7]: [8]: [9]: [10]: [11]:	Additionally in coalize to compare the heights of the bars than in its compare the size of pip pars. The numbers or they was make in possible of read the acute survey in probability of seed data, as it hope to be made in the face with probability and should five processing them in the data of the probability of the data and that a parketyror of the data can the data parketyror of the data can be data and the data of the data can the data can be data and the data can be data and the data can be data and the data can be data data and the data data data data data data data dat
[8]: [7]: [10]: [11]: [12]: [13]:	Additionally is easier to compare the heights of the basis born is to compare the stars of pic parts. The numbers of the year mode in possible or made the actual charactery probability of social data, and by bosing or of received and the actual probability of social data. A part of the Start of the Star