

# Project Report on Data Science in manufacturing industry



**Tata Technologies Ltd,  
Thane, India**

**12 June 2017 - 11 August 2017**

**By Aditya Oak**

**Under the guidance of: Mr. Sachidanand Tripathi**

**TATA TECHNOLOGIES LTD., THANE, INDIA**

**CERTIFICATE**

This is to certify that the project titled:

**Data Science for Manufacturing Industry**

Submitted by

**Aditya Oak**

is a bonafide work carried out by the above mentioned intern under the supervision of  
**Mr. Sachidanand Tripathi** and it is submitted towards the partial fulfilment of the  
requirement of Project Training

Mr. Sachidanand Tripathi  
Project Mentor

Date:

Place:

## Acknowledgements:

I would like to take this opportunity to thank all the people who helped me during the course of this internship. The successful completion of projects would not have been possible without help from the Data Analytics team of Tata Technologies Ltd.

Firstly, I would like to thank Mr Rajendra Borkar who recommended my profile for this internship. I was able to harness immense learning outcomes

Special thanks to my mentors Mr. Kamlesh Pandey and Mr. Sachidanand Tripathi for guiding me through. They continuously focused on putting my expertise and knowledge of Mechanical Engineering to good use. Especially, the data architecture for overhead cranes and mechanical press machines.

Being new to the field, I learnt a lot from the likes of data analytics team members including Hrishikesh Kherdekar, Aniket Parulekar and Pushkar Lanke. They provided continuous support in programming and analytical concepts and with silliest of my doubts.

I would also like to extend my gratitude towards HR & Admin for the warm welcome and maintaining a healthy work environment.

TATA  
TECHNOLOGIES

## Table of Contents:

Introduction

Data Analysis and Data Science

Learning Python

Overhead crane project:

- Defining use case
- Defining attributes

Mechanical Press Project:

- Defining use case
- Defining attributes

Machine Learning on Titanic Dataset

- Dataset and source
- Use case
- Attributes
- Graphical analysis

Mercedes Benz Dataset

- One-hot encoding
- Target variable (y)
- Unique values

TATA

TECHNOLOGIES

## Introduction:

This internship was primarily focused on the aspects of Data science, which is currently an upcoming field and in great demand. It started off with a brief insight into data analysis and its requirement in today's world. It then headed to an introduction to Data Science which gives an idea about how data can be used to make business decisions.

We then step out on the code playground where we explore the programming capacities of Python, both basic and advance and its role in the in the world of Data.

My primary education in the Mechanical stream and previous work in the crane industry helped me narrow down to the overhead crane project. List of attributes, its source. Thereon to the press machine project. List of attributes, source, etc.

Analysis of titanic data set to predict who survived based on attributes such as Sex, Age, Travel class. It included Data exploration, defining use case and graphical analysis of the dataset.

Mercedes Benz green manufacturing data set. Immense learning opportunity from team associated with the competition. It included on-hot encoding, target variable and unique values.

IMA  
TECHNOLOGIES

## Data Analysis and Data Science

In this part, I have started my journey towards data analysis and its aspects through Udacity's Intro to Data Analysis and Data Science

Throughout the course I learnt that Data Analysis is essentially understanding the stories that the data is trying to narrate, over a period of time. These stories give insights into the past and its impact on the present and future scenarios, more commonly known as predictive analysis. To interpret this, we ought to adapt a logical reasoning approach and a good amount of stats 101.

Data analytics seeks to provide operational observations into issues that we either know we know or know we don't know. Descriptive analytics, for example, quantitatively describes the main features of a collection of data. Predictive analytics, that focus on correlative analysis, predicts relationships between known random variables or sets of data in order to identify how an event will occur in the future. For example, identifying the where to sell personal power generators and the store locations as a function of future weather conditions (e.g., storms). While the weather may not have caused the buying behavior, it often strongly correlates to future sales.

**TATA**  
**TECHNOLOGIES**

## Python Programming

An important aspect of Data Science is the ability to code in languages such as R, Python, etc. I had experience working on Java since my high school and that helped me dive right into the advance concepts of Python, which is very similar to Java.

Python is largely used in the field of data science and analytics. Its ability to read csv files and perform operations on it, without opening the file itself, gives Python a strong edge in data science. I started off by getting familiar with how Python works differently than Java and then built the advanced concepts over it.

### Pandas:

I had worked with pandas on the 2 online courses that I did on Udacity. It showcased the true power of Python. I used `read_csv` extensively to import csv files.

`read_csv()`: Read CSV (comma-separated) file into DataFrame.

Following are some commonly used pandas series functions:

- `Series.values()`: returns Series as ndarray or ndarray-like depending on the dtype
- `Series.size()`: returns the number of elements in the underlying data
- `Series.isnull()`: returns a boolean same-sized object indicating if the values are null.
- `Series.groupby()`: group series using mapper (dict or key function, apply given function to group, return result as series) or by a series of columns.
- `Series.describe()`: Generates descriptive statistics that summarize the central tendency, dispersion and shape of a dataset's distribution, excluding NaN values
- `Series.unique()`: Return unique values in the object.
- `Series.nunique()`: Return number of unique elements in the object.
- `Series.is_unique()`: Return boolean if values in the object are unique

Pandas also have a dataframe which supports values entered in row x column format. Some of the functions of a pandas DataFrame are:

- `DataFrame.head([n])`: Returns first n rows
- `DataFrame.all()`: Return whether all elements are True over requested axis
- `DataFrame.any()`: Return whether any element is True over requested axis
- `DataFrame.count()`: Return Series with number of non-NA/null observations over requested axis.
- `DataFrame.plot.area()`: Area plot
- `DataFrame.plot.bar()`: Vertical bar plot
- `DataFrame.plot.barh()`: Horizontal bar plot
- `DataFrame.plot.box()`: Boxplot
- `DataFrame.plot.density(**kws)`: Kernel Density Estimate plot
- `DataFrame.plot.hexbin()`: Hexbin plot
- `DataFrame.plot.hist()`: Histogram
- `DataFrame.plot.kde(**kws)`: Kernel Density Estimate plot
- `DataFrame.plot.line()`: Line plot
- `DataFrame.plot.pie()`: Pie chart
- `DataFrame.plot.scatter()`: Scatter plot

- `DataFrame.boxplot()`: Make a box plot from DataFrame column optionally grouped by some columns or
- `DataFrame.hist()`: Draw histogram of the DataFrame's series using matplotlib / pylab.

**TATA**  
**TECHNOLOGIES**



## Overhead Crane data points:

With the results that data driven technology delivers, it is only a matter of time that all industries worldwide will turn to data science to make better decisions for their businesses. Manufacturing industry, in particular, produces large amounts of data about their equipments, processes and final products. Consequently, analyzing this data to understand the production efficiency of the business becomes paramount.

I, along with inputs from my mentor, decided to take up the material handling industry (which I briefly worked for) and explore the data points that operate under it. Material handling industry has been around for quite some time, ever since engineers started using pulleys, ropes, gears, etc. It is broadly divided into fixed (overhead cranes, wall crawler, jib crane and gantry crane) and mobile (which can move about like a vehicle). For our data points, we consider only fixed type and in it we focus mainly on overhead cranes.

Overhead cranes are mainly found in manufacturing plants where parts weighing over a ton are required to move around for assembly. Following are some basic parts of an overhead crane:

**Girders:** The main structural body of the crane which spans between two ends of the plant and travels along the length of the plant. Depending on its load carrying capacity, it is further divided into single girder and double girder crane.

**Hoist:** The trolley like structure which is placed on the girder and is free to move about the length of the girder. This part has the main hoisting mechanism which carries out the material handling operation

**Lower block assembly:** As the name suggests, this is a lower block which hangs about from the hoist and has pulleys, sheaves and wire rope as part of it. Also, the hook is attached at the bottom of this.

**Hook:** Probably the most commonly known crane part on which the load is lifted. It is kept in a state of free rotations from the lower block assembly and is made by forging metal alloys.

**Defining use case:** It was important that we define what we are trying to predict with this analysis. After considering various aspects of the production time, it was decided to predict whether the crane will breakdown or not. This was crucial since the crane downtime had a significant impact on the overall production time.

**Defining attributes:** An overhead crane has numerous characteristics and conditions which work together to provide effective material handling solutions. My primary education in Mechanical Engineering was going to come in handy through this. Following is the list of attributes and their sub-attributes:

### **Lubrication:**

Part to be lubricated - gears, bearings, wheels, hooks, blocks, electrical parts

Method of application- Splash, Manual, Grease Nipple, Grease Pump, Felt lubrication

Interval- A – Change after 2000 hrs. of operation or 6 months

B – Change after 1500 hrs. of operation or 3 months

C – Change after 1000 hrs. of operation or 3 months

D – Change after 750 hrs. of operation or 2 months

E – Apply weekly or every 50 hrs.

F – Apply daily or every 8 hrs.

G – Apply monthly or every 50 hrs.

**Wear:**

Part – Gears, Wheels, Sheaves, Brake lining, Wire rope  
Wear - % of part size

**Brakes:**

High speed travel  
Stopping distance - % of high speed travel  
Drift – Yes / No

**Controls:**

Support against strain	S- Satisfactory, U- Unsatisfactory, N- Not applicable
Identification/la bels	S- Satisfactory, U- Unsatisfactory, N- Not applicable
Warning labels	S- Satisfactory, U- Unsatisfactory, N- Not applicable
General condition	S- Satisfactory, U- Unsatisfactory, N- Not applicable

**Hook:**

Safety Latch – S,U,N  
Deformation – S,U,N  
Size of hook  
Wear – S,U, N OR Percentage of size of hook (<10%)  
Cracks, nicks, gouges – S,U,N  
Attachment points – S,U,N  
Self-locking operation – S,U,N  
Bending – Y / N  
Twisting – Y/N  
Cracks – Y/N  
Safety Latch in place – Y/N  
Hook free rotation (grinding or sound) – Y/N

Hook drift (Up) – Y/N  
Hook drift (Down) –Y/N

**Lower block assembly:**

Structural Damage –Y/N  
Cracks in any component –Y/N  
Capacity Markings present –Y/N  
Sheaves rotate freely (grinding or sound) – Y/N  
Sheaves are smooth –Y/N  
Sheave guards intact – Y/N

**\*Chain: (\*If not wire rope, but not both)**

Operation – S,U,N  
Lubrication – S,U,N  
Defects/wear –S,U,N

**Wire rope**

Distortion – S,U,N  
Corrosion – S,U,N  
Broken wires/cuts – S,U,N  
Reeving – S,U,N  
% Reduction in diameter – (not more than 33%)  
Broken wires – Y/N  
Kinking – Y/N  
Cutting – Y/N  
Crushing – Y/N  
Unstranding –Y/N  
Thermal Damage – Y/N

**Structure:**

Load rating marked – S,U,N  
Defects, Cracks, Welds – S, U,N  
Air System – S, U,N  
Hydraulic System – S, U,N

**Bridge:**

Wear/defects – S,U,N

**Trolley:**

Wear/defects – S,U,N

**Hoist:**

Wear/defects - S,U,N

**Electrical:**

Emergency Disconnect – S,U,N

Identification Labels – S,U,N

Warning labels – S,U,N

**Function:**

Limit Switches - S,U,N

Hoist Mechanisms – S,U,N

Bridge Mechanisms – S,U,N

Trolley Mechanisms – S,U,N

**Cab:**

Fire extinguisher –S,U,N

Housekeeping –S,U,N

Lightning – S,U,N

Hand Signal Chart – S,U,N

**Crane inspector:**

Field hours of experience – hours (should be ideally >2000)

**Girder:**

Span (L) – ft (')

Increase in span (A) – inches (")

Rail to rail elevation (D) – inches (")

Straightness (B) – inches (")

Elevation (C) – inches (")

**Gears:**

Grinding sound – Y/N

Axial shift – Y/N

Oil change - Y/N

Hours of operation – hours (ideally  $\leq 150$  hours)

**Travelling wheels:**

Cracks – Y/N

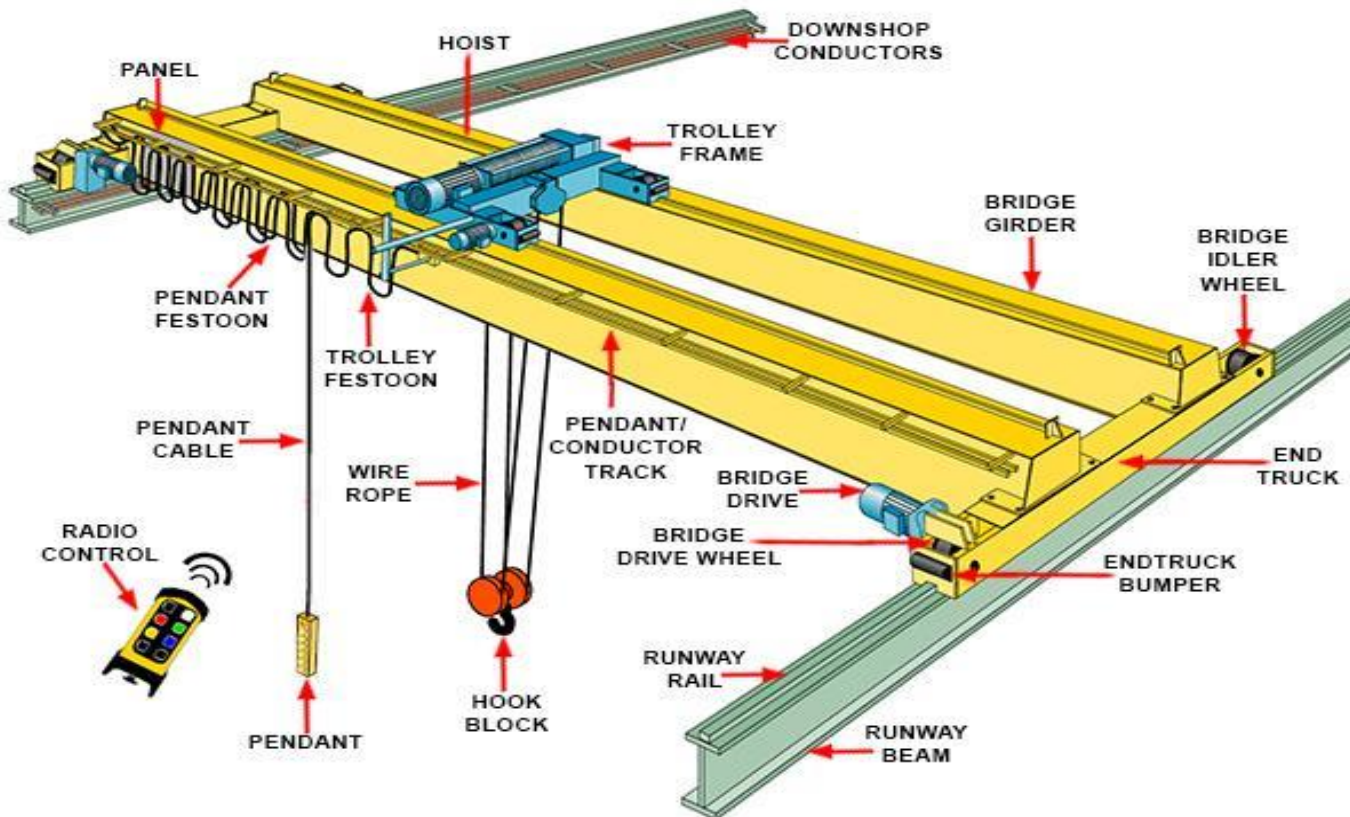
Worn out collar rims –Y/N

## Indicators:

**Leading indicators:** intrafocus.com defines leading indicators as a predictive measurement which can influence change for the future, for example; the percentage of people wearing hard hats on a building site is a leading safety indicator. It further adds, lead indicators are always more difficult to determine .They are predictive and therefore do not provide a guarantee of success. This not only makes it difficult to decide which lead indicators to use, it also tends to cause heated debate as to the validity of the measure at all.

**Lagging indicators:** intrafocus.com defines a lagging indicator as an output measurement, for example; the number of accidents on a building site is a lagging safety indicator. The difference between the two is a leading indicator and a lagging indicator can only record what has happened.

A closer look at the problem revealed the need to determine leading and lagging indicators. Thus, operational hours of crane, fatigue cycle, maintenance log, daily inspection log were the **leading indicators** and crane accidents in the past, crane breakdown history, crane idle time, time delay due to breakdowns, were the **lagging indicators**.



## Mechanical Press data points:

Applying Data Science to the manufacturing industry meant starting with such an equipment that is present universally in most, if not all, industries. Sachidanand sir came up with an idea which was the best fit for this exercise: the mechanical press machine. Furthermore, it was essential for these machines that their operational data be capturable.

Wikipedia defines press machines a machine tool that changes the shape of a workpiece by the application of pressure. Press machines are largely used in industries where metal is required in specific shapes. Depending on the malleability of metals, the operation of press machines differ. Furthermore, the method of application of pressure to the workpiece also determines whether it is a mechanical or a hydraulic press

**Mechanical press** uses the traditional reciprocating motion of connecting rod and crankshaft to deliver pressure strokes to the workpiece. It is usually driven by a power supply which actuates the crankshaft and sets the process in action.

**Hydraulic press**, on the other hand, uses pressure from fluids to carry out the reciprocating motion. When liquid flows in the cylinder, it pushes the piston out of it. This takes the ram on the horizontal bed. At the bottom dead center of the stroke, the liquid is pulled back, which causes the piston to retract to its original position. Thus, the reciprocating motion continues.

### Attributes:

**Press type** – Mechanical, Hydraulic, and Pneumatic

**Crank Shaft Diameter** – mm

**Stroke adjustment** – mm

**Slide adjustment** – mm

**Hole in Ram** – mm

**Hole in bed** – mm

**Area of bed** – mm<sup>2</sup>

**Bed to Ram distance** – mm

**Horse power** – HP

**No. of rotations** – RPM

**Weight of machine - kg**

**Clutch type** – Full revolution clutch, part revolution clutch (more hazardous)

**Control type** – Manual (foot pedal/palm switches), Automatic

**Hazard time(T)** – seconds

*Presses with part-revolution clutches, the hazard time is defined as the stopping time of the press ram.  
On presses with full-revolution clutches, the hazard time is defined as the maximum possible time required for the ram to complete one down stroke.*

**Safety Distance (D)** - the minimum safe distance between the point of operation and the palm-button.  
(Ideally  $1.6 \times T$ ) (Co-relates with Hazard Time)

**Number of surge tanks - nos.**

**Surge tanks drained** – Y/N

**Number of air pressure switches** – nos.

**Brake lining worn** – Y/N

**Excessive grease on brake lining** – Y/N/Not known

**Brake makes excessive noise** – Y/N

**Electrical Wiring properly installed:** Y/N

**Motor start button protected against accidental operation** –Y/N

**Drive motor start (magnetic type)** –Y/N

**Anti-repeat** – Y/N

**Coating of Boron Nitride powder on billets** – Y/N

**Float on container** – Y/N

**Oil leaks** – Y/N

**Burp cycle operating correctly** – Y/N (The function of the burp cycle is to expel any entrapped air between the front section of the container and the die, preventing air bubbles and blisters forming in the extrusion.)

**Pump PF3 pressure - PSI** (should be between 500 to 600 psi)

**Pump PF3A pressure – PSI** (should be >=150 psi)

**Pump PF4A pressure –** (should be between 500 to 600 psi)

**Blade and die face clearance – mm** (ideally 0.12-2.5 mm)

**Lubrication between butt and blade – Y/N**

**Butt shear in full up position – Y/N**

**Load valve energized – Y/N**

**Press condition – Working/Not working**

Sample readings:

A rough look at how the press data correlates to each other is shown below. The part dimensions vary significantly with an increase in the machine's tonnage.

								Unit : mm		
MODEL	5 Ton	10 Ton	20 Ton	30 Ton	50 Ton	80 Ton	100 Ton	150 Ton	200 Ton	250 Ton
<b>Crank Shaft Dia</b>	50	58	73	83	95	114	127	152	158	165
<b>Stroke Adjustment</b>	6 TO 25	6 TO 50	10 TO 62	13 TO 75	13 TO 100	13 TO 112	13 TO12 5	13 TO 125	165	165
<b>Slide Adjustment</b>	30	40	50	50	50	60	60	60	70	70
<b>Hole In Ram</b>	19	25	32	38	51	51	55	60	63	63
<b>Hole In Bed</b>	51	70	89	102	127	127	191	200	216	229
<b>Length &amp; Width of Bed</b>	230 X 142	381 X 288	455 X 250	508 X 355	650 X 400	750 X 500	800 X 650	800 X 650	1016 X 736	1067 X 778



<b>Dis To Bed To Ram</b>	150	203	230	266	350	450	450	450	519	544
<b>H.P/r.p.m</b>	0.75 / 1440	1 / 1440	2 / 1440	3 / 1440	5 / 1440	7.5 / 1440	10 / 1440	15 / 1440	20 / 1440	25 / 1440
<b>Weight Of Approx Kg.</b>	255	525	1100	1300	2300	3000	4500	5500	9100	11700



Muller press machine at work

# Machine Learning from Titanic

## Dataset and source

A classic example of machine learning as defined by kaggle is learning from disaster. The RMS Titanic which sunk to the bottom of the ocean, killing a large number of passengers onboard, is an excellent example of how a person's characteristics define their survival. I took up this data from a past data competition on kaggle.com. This whole idea of a big picture perspective and studying who survived the disaster, is indeed an intriguing one for an aspiring data scientist.

## Use case:

With the dataset at hand, the target variable was obvious: who survived? This was further divided into what groups of people survived. Or what combination of characteristics of a passenger ensured their survival.

## Attributes:

*PassengerId* - primary key for passengers

*Survived* - Binary (0- Not survived, 1 - Survived)

*Pclass* - Ticket class of passenger (1,2,3)

*Sex* - Binary (Male/Female)

*Age* - Passenger age

*SibSp* - No. of siblings / spouses aboard

*Parch* - No. of parents / children aboard

*Ticket* - Ticket number

*Fare* - ticket fare

*Cabin* - Cabin number of passenger

*Embarked* - Port of embarkation (C,Q,S)

## Data exploration

I used Python 2 to run analysis of the dataset.

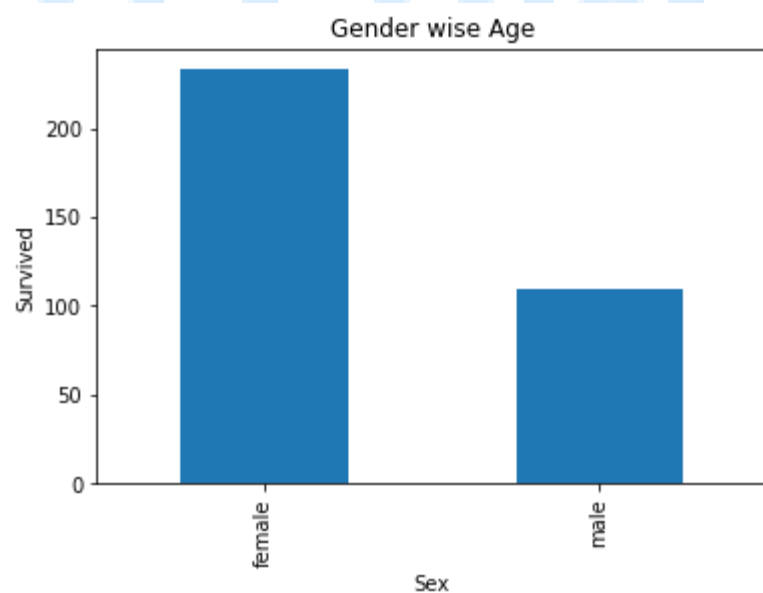
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

Death count: **549**

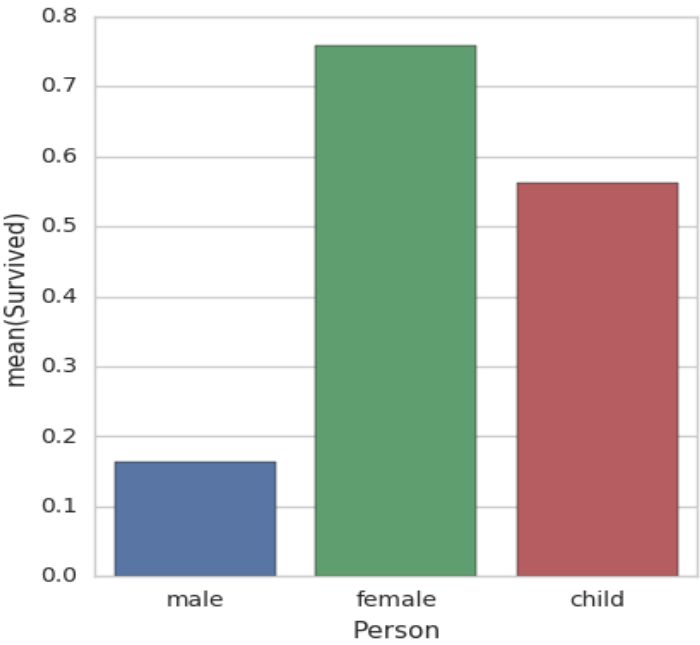
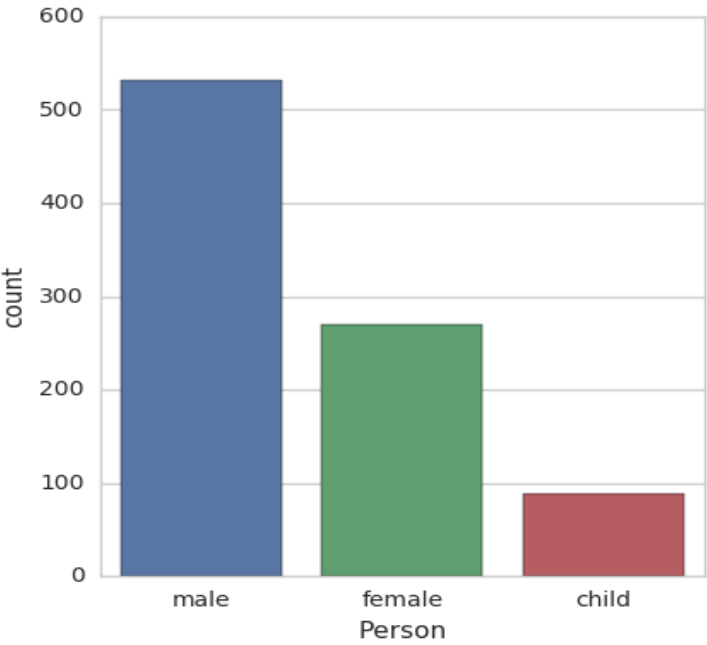
Survived count: **342**

	Sex	Survived
1	male	109
0	female	233

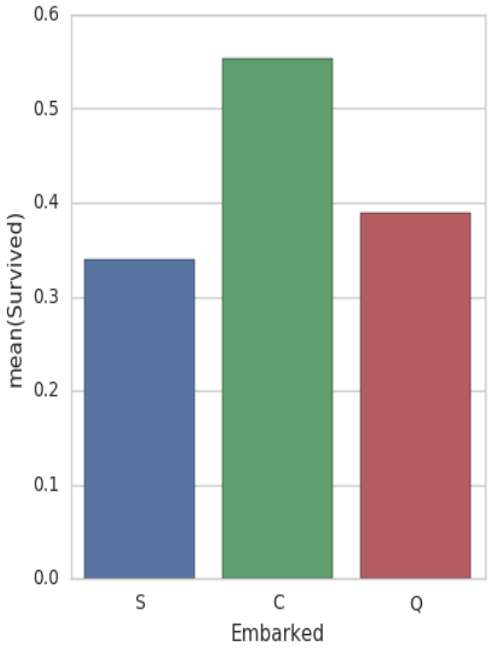
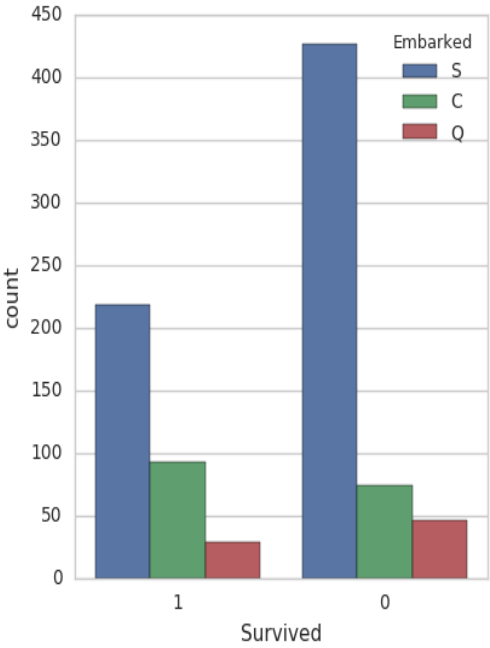
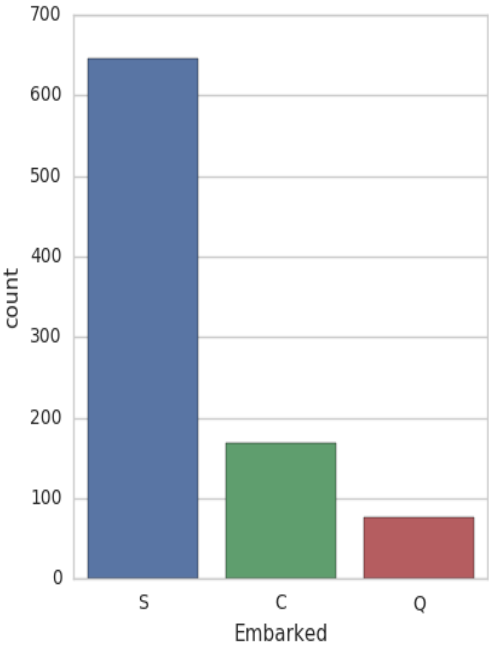
## Graphical analysis



Children(age < ~16) on aboard seem to have a high chances for Survival. So, we can classify passengers as males, females, and child



Embarkation:



## Mercedes Benz dataset:

I had the opportunity to learn intermediate concepts of data analytics from the team working on the Mercedes Benz green manufacturing competition on kaggle. As a beginner in Data Science, I traced back the initial few steps that they carried out in order to have exposure to the new concepts.

**Dataset:** As per kaggle, This dataset contains an anonymized set of variables, each representing a custom feature in a Mercedes car. For example, a variable could be 4WD, added air suspension, or a head-up display. The ground truth is labeled 'y' and represents the time (in seconds) that the car took to pass testing for each variable.

### One-Hot encoding:

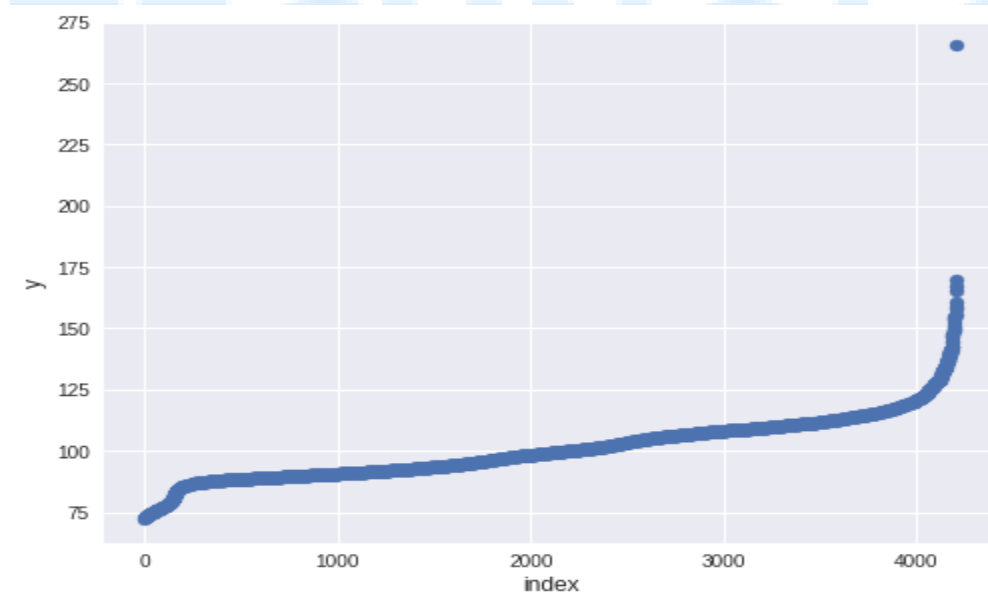
[machinelearningmastery.com](https://machinelearningmastery.com) defines one hot encoding as a representation of categorical variables as binary vectors.

This first requires that the categorical values be mapped to integer values.

Then, each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1.

In this case, we had attributes X0, X1, X2, X3, X4, X5, X6, X8 whose values were categorical. These needed to be converted to numeric data so that machine learning algorithms can work on it.

### Target variable (y):



As seen above, the target variable has one outlier but all other values are closely associated.

## Unique values:

If a column contains only 1 unique value, the column can be dropped since it won't influence the target variable in any way. However, if there are more than 1 unique values in the column, it is bound to have a significant impact on the behaviour of the target variable. Thus, it is important to distinguish such columns.

### Columns containing the unique values : [0, 1]

['X10', 'X12', 'X13', 'X14', 'X15', 'X16', 'X17', 'X18', 'X19', 'X20', 'X21', 'X22', 'X23', 'X24', 'X26', 'X27', 'X28', 'X29', 'X30', 'X31', 'X32', 'X33', 'X34', 'X35', 'X36', 'X37', 'X38', 'X39', 'X40', 'X41', 'X42', 'X43', 'X44', 'X45', 'X46', 'X47', 'X48', 'X49', 'X50', 'X51', 'X52', 'X53', 'X54', 'X55', 'X56', 'X57', 'X58', 'X59', 'X60', 'X61', 'X62', 'X63', 'X64', 'X65', 'X66', 'X67', 'X68', 'X69', 'X70', 'X71', 'X73', 'X74', 'X75', 'X76', 'X77', 'X78', 'X79', 'X80', 'X81', 'X82', 'X83', 'X84', 'X85', 'X86', 'X87', 'X88', 'X89', 'X90', 'X91', 'X92', 'X94', 'X95', 'X96', 'X97', 'X98', 'X99', 'X100', 'X101', 'X102', 'X103', 'X104', 'X105', 'X106', 'X108', 'X109', 'X110', 'X111', 'X112', 'X113', 'X114', 'X115', 'X116', 'X117', 'X118', 'X119', 'X120', 'X122', 'X123', 'X124', 'X125', 'X126', 'X127', 'X128', 'X129', 'X130', 'X131', 'X132', 'X133', 'X134', 'X135', 'X136', 'X137', 'X138', 'X139', 'X140', 'X141', 'X142', 'X143', 'X144', 'X145', 'X146', 'X147', 'X148', 'X150', 'X151', 'X152', 'X153', 'X154', 'X155', 'X156', 'X157', 'X158', 'X159', 'X160', 'X161', 'X162', 'X163', 'X164', 'X165', 'X166', 'X167', 'X168', 'X169', 'X170', 'X171', 'X172', 'X173', 'X174', 'X175', 'X176', 'X177', 'X178', 'X179', 'X180', 'X181', 'X182', 'X183', 'X184', 'X185', 'X186', 'X187', 'X189', 'X190', 'X191', 'X192', 'X194', 'X195', 'X196', 'X197', 'X198', 'X199', 'X200', 'X201', 'X202', 'X203', 'X204', 'X205', 'X206', 'X207', 'X208', 'X209', 'X210', 'X211', 'X212', 'X213', 'X214', 'X215', 'X216', 'X217', 'X218', 'X219', 'X220', 'X221', 'X222', 'X223', 'X224', 'X225', 'X226', 'X227', 'X228', 'X229', 'X230', 'X231', 'X232', 'X234', 'X236', 'X237', 'X238', 'X239', 'X240', 'X241', 'X242', 'X243', 'X244', 'X245', 'X246', 'X247', 'X248', 'X249', 'X250', 'X251', 'X252', 'X253', 'X254', 'X255', 'X256', 'X257', 'X258', 'X259', 'X260', 'X261', 'X262', 'X263', 'X264', 'X265', 'X266', 'X267', 'X269', 'X270', 'X271', 'X272', 'X273', 'X274', 'X275', 'X276', 'X277', 'X278', 'X279', 'X280', 'X281', 'X282', 'X283', 'X284', 'X285', 'X286', 'X287', 'X288', 'X291', 'X292', 'X294', 'X295', 'X296', 'X298', 'X299', 'X300', 'X301', 'X302', 'X304', 'X305', 'X306', 'X307', 'X308', 'X309', 'X310', 'X311', 'X312', 'X313', 'X314', 'X315', 'X316', 'X317', 'X318', 'X319', 'X320', 'X321', 'X322', 'X323', 'X324', 'X325', 'X326', 'X327', 'X328', 'X329', 'X331', 'X332', 'X333', 'X334', 'X335', 'X336', 'X337', 'X338', 'X339', 'X340', 'X341', 'X342', 'X343', 'X344', 'X345', 'X346', 'X348', 'X349', 'X350', 'X351', 'X352', 'X353', 'X354', 'X355', 'X356', 'X357', 'X358', 'X359', 'X360', 'X361', 'X362', 'X363', 'X364', 'X365', 'X366', 'X367', 'X368', 'X369', 'X370', 'X371', 'X372', 'X373', 'X374', 'X375', 'X376', 'X377', 'X378', 'X379', 'X380', 'X382', 'X383', 'X384', 'X385']

### Columns containing the unique values : [0]

['X11', 'X93', 'X107', 'X233', 'X235', 'X268', 'X289', 'X290', 'X293', 'X297', 'X330', 'X347']

## References

- <https://classroom.udacity.com/courses/ud170>
- <https://classroom.udacity.com/courses/ud359>
- kaggle.com
- machinelearningmastery.com
- CICB Overhead Crane/Hoist Monthly Checklist
- Cranex maintenance manual
- Michigan Occupational Safety and Health Administration (MIOSHA)
- CCAA Overhead Crane Inspection Checklist
- [www.stlcrane.com](http://www.stlcrane.com)
- www.castool.com

TATA  
TECHNOLOGIES