Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

PROJECT REPORT

ON

Engineering graduate salary prediction

Submitted by

TANVI KUMARI

11912593

Programme: Bachelor of Technology

Section: KM006

Course Code: INT217

Under the Guidance of

KOMAL ARORA :17783

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara
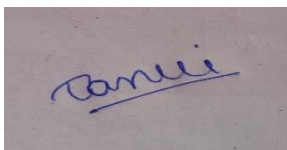
<u>DECLARATION</u>

I, Tanvi Kumari, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/01/2021

Tanvi Kumari

Registration No: 11912593

Signature :

<u>ACKNOWLEDGEMENT</u>

TABLE OF CONTENTS

INTRODUCTION

➢ Data management is important because the data your organization creates is a very valuable resource.

➢ The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.

➢ In that case, you would then have to spend time and resources again to get that same business intelligence you already had.

➢ And on that data analysis is carried out which show visualization of our problems in efficient way.

➢ Data Analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

➢ This project is based on such data analysis on Engineering graduate salary prediction.

➢ The salary and the jobs engineers are offered right after graduation. Various factors such as college grades, candidate skills, the proximity of the college to industrial hubs, the specialization one have market conditions for specific industries determine this. On the basis of these various factors, main objective is to determine the salary of an engineering graduate in India.

➢ The dataset contains 37 data fields.

OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:

- Highest Salary on the basis of specification.

- Salary based on GPA

- Salary based on LOGICAL

- Exploration on 12th Board

- Exploration on 12th Board

- Exploration on Degree

SOURCE OF DATASET:

Source of dataset: https://www.kaggle.com/manishkc06/engineering-graduate-salary-prediction

The dataset is based on more than 1000 candidates record and their salary after graduating. The columns included in the dataset are given below:

- ID:
    A unique ID to identify a candidate
- Salary:
    Annual CTC offered to the candidate
- Gender:
    Candidate's gender
- DOB:
    Date of birth of the candidate
- 10percentage:
    Overall marks obtained in grade 10 examinations
- 10board:
    The school board whose curriculum the candidate followed in grade 10
- 12graduation:
    Year of graduation - senior year high school
    12percentage:
    Overall marks obtained in grade 12 examinations
- 12board:
    The school board whose curriculum the candidate followed
- College ID:
    Unique ID identifying the university/college which the candidate attended for her/his undergraduate
- College Tier:
    Each college has been annotated as 1 or 2. The annotations have been computed from the average scores obtained by the students in the college/university. Colleges with an average score above a threshold are tagged as 1 and others as 2.
- Degree:
    Degree obtained/pursued by the candidate
- Specialization:
    Specialization pursued by the candidate
- College GPA:
    Aggregate GPA at graduation
- College City ID:
    A unique ID to identify the city in which the college is located in.
- College City Tier:
    The tier of the city in which the college is located in. This is annotated based on the population of the cities.

7

- College State:

  Name of the state in which the college is located
- Graduation Year:

  Year of graduation (Bachelor's degree)
- English:

  Scores in English section
- Logical:

  Score in Logical ability section
- Quant:

  Score in Quantitative ability section
- Domain:

  Scores in domain module
- Computer Programming:

  Score in Computer programming section
- Electronics And Semicon:

  Score in Electronics & Semiconductor Engineering section
- Computer Science:

  Score in Computer Science section
- Mechanical Engineering:

  Score in Mechanical Engineering section
- Electrical Engineering:

  Score in Electrical Engineering section
- Electrical Engineering:

  Score in Electrical Engineering section
- Telecom Engineering:

  Score in Telecommunication Engineering section
- Civil Engineering:

  Score in Civil Engineering section
- conscientiousness:

  Scores in one of the Sections of personality test
- agreeableness:

  Scores in one of the sections of personality test
- extraversion:

  Scores in one of the sections of personality test
- nueroticism:

  Scores in one of the sections of Personality test
- openness *to* experience:

  Scores in one of the sections of personality test

Sample of dataset with data fields is given below:

ID	Gender	DOB	10percent	10board	12graduat	12percent	12board	CollegeID	CollegeTi	Degree	Specializa	collegeGP	CollegeCi	CollegeCi	CollegeSt	Graduatio	English	Logical	Quant
604399	f	22-10-1990	87.8	cbse	2009	84	cbse	6920	1	B.Tech/B.	instrumer	73.82	6920	1	Delhi	2013	650	665	810
988334	m	15-05-1990	57	cbse	2010	64.5	cbse	6624	2	B.Tech/B.	computer	65	6624	0	Uttar Prad	2014	440	435	210
301647	m	21-08-1989	77.33	maharash	2007	85.17	amravati c	9084	2	B.Tech/B.	electronic	61.94	9084	0	Maharash	2011	485	475	505
582313	m	04-05-1991	84.3	cbse	2009	86	cbse	8195	1	B.Tech/B.	computer	80.4	8195	1	Delhi	2013	675	620	635
339001	f	30-10-1990	82	cbse	2008	75	cbse	4889	2	B.Tech/B.	biotechno	64.3	4889	1	Tamil Nad	2012	575	495	365
609356	f	02-12-1989	83.16	icse	2007	77	cbse	10950	1	M.Tech./N	instrumer	99.93	10950	0	Punjab	2013	535	595	620
1081649	f	17-04-1989	72.5	state boar	2007	53.2	state boar	14381	2	B.Tech/B.	mechanic	68	14381	1	West Ben	2013	510	495	405
610842	f	11-04-1991	77	state boar	2009	88	state boar	13208	2	B.Tech/B.	computer	71	13208	1	Telangana	2013	370	470	280
1183070	m	25-11-1992	76.8	state boar	2010	87.7	state boar	5338	2	B.Tech/B.	informatic	73.15	5338	0	Andhra Pr	2014	510	555	440
794062	f	15-03-1993	57	state boar	2009	73	state boar	8346	2	B.Tech/B.	computer	70.08	8346	0	Uttar Prad	2014	500	410	560
1088206	m	21-06-1990	77	state boar	2008	75	state boar	13424	2	B.Tech/B.	electronic	62	13424	0	Maharash	2013	675	630	485
1279958	m	02-07-1992	81.2	state boar	2008	79.9	state boar	64	2	B.Tech/B.	instrumer	67.67	64	0	Uttar Prad	2013	395	565	645
471413	f	24-12-1991	85	delhi boar	2009	88	all india b	57	2	B.Tech/B.	informatic	85	57	0	Haryana	2013	495	445	605
1088423	f	09-08-1992	90	state boar	2009	82.1	state boar	2998	2	B.Tech/B.	computer	85	2998	0	Telangana	2014	640	530	705
1066680	m	12-09-1991	86.4	cbse	2009	86.2	cbse	1906	2	B.Tech/B.	electronic	81.4	1906	1	West Ben	2014	720	630	750
407672	m	30-09-1990	84.13	0	2008	77	0	3801	2	B.Tech/B.	informatic	75.2	3801	0	Karnataka	2012	385	515	465
205633	m	25-05-1988	81.7	hse	2005	75.8	chse	5508	2	B.Tech/B.	computer	78.7	5508	0	Orissa	2011	475	245	485
924541	f	25-02-1992	86	cbse	2010	89	cbse	429	2	B.Tech/B.	computer	73.9	429	0	Uttar Prad	2014	520	570	620
512353	m	25-10-1991	66.15	state boar	2009	54	state boar	3603	2	B.Tech/B.	computer	66	3603	0	Maharash	2013	360	495	280
1136577	m	22-05-1991	79.29	icse	2009	68.67	cbse	5298	2	B.Tech/B.	computer	76	5298	0	Tamil Nad	2013	545	505	595
781327	m	21-11-1987	60	state boar	2006	50	state boar	11013	2	B.Tech/B.	computer	69.94	11013	0	Orissa	2013	240	340	430
764522	f	14-01-1992	58.4	cbse	2008	64.8	cbse	85	2	MCA	computer	80	85	0	Haryana	2014	440	415	554
203323	m	03-11-1986	61	0	2003	59	0	3716	2	B.Tech/B.	informatic	63	3716	1	Karnataka	2011	355	565	495
267354	m	05-08-1988	50	board of s	2008	64	state boar	7564	2	B.Tech/B.	electronic	70	7564	0	Haryana	2011	345	365	445
839626	f	10-04-1991	67.06	state boar	2008	70.67	state boar	13271	2	B.Tech/B.	computer	55.5	13271	1	Maharash	2012	440	440	385
1169332	f	09-03-1990	67	cbse	2008	61	cbse	16664	2	B.Tech/B.	computer	75	16664	0	Chhattisga	2012	370	410	320

CollegeCi	CollegeSt	Graduatio	English	Logical	Quant	Domain	Computer	Electronic	Computer	Mechanic	ElectricalE	TelecomE	CivilEngg	conscienti	agreeable	extraversi	nueroticis	openess_	Salary
1	Delhi	2013	650	665	810	0.694479	485	366	-1	-1	-1	-1	-1	-0.159	0.3789	1.2396	0.1459	0.2889	445000
0	Uttar Prad	2014	440	435	210	0.342315	365	-1	-1	-1	-1	-1	-1	1.1336	0.0459	1.2396	0.5262	-0.2859	110000
0	Maharash	2011	485	475	505	0.824666	-1	400	-1	-1	-1	260	-1	0.51	-0.1232	1.5428	-0.2902	-0.2875	255000
1	Delhi	2013	675	620	635	0.990009	655	-1	-1	-1	-1	-1	-1	-0.4463	0.2124	0.3174	0.2727	0.4805	420000
1	Tamil Nad	2012	575	495	365	0.278457	315	-1	-1	-1	-1	-1	-1	-1.4992	-0.7473	-1.0697	0.06223	0.1864	200000
0	Punjab	2013	535	595	620	0.37606	455	300	-1	-1	-1	313	-1	0.8463	-0.6201	-0.7585	-0.995	-0.2859	440000
1	West Ben	2013	510	495	405	0.829585	-1	-1	-1	469	-1	-1	-1	0.1282	-0.4536	0.3174	0.9066	-0.0943	150000
1	Telangana	2013	370	470	280	0.70409	465	-1	-1	-1	-1	-1	-1	0.1282	0.5454	0.4711	0.9066	-0.2859	105000
0	Andhra Pr	2014	510	555	440	0.744758	525	-1	438	-1	-1	-1	-1	-0.159	-0.4536	0.1637	0.5262	-0.0943	195000
0	Uttar Prad	2014	500	410	560	0.622643	385	-1	407	-1	-1	-1	-1	-0.7335	-0.4536	-0.2974	1.4136	0.6721	200000
0	Maharash	2013	675	630	485	0.207392	405	260	-1	-1	-1	-1	-1	1.7081	0.5454	0.9322	-0.4879	0.8637	335000
0	Uttar Prad	2013	395	565	645	-1	495	-1	376	-1	-1	-1	-1	0.7027	0.7119	0.01	-0.995	0.6721	300000
0	Haryana	2013	495	445	605	0.765674	485	-1	-1	-1	-1	-1	-1	-1.9629	-2.1903	-2.775	0.4148	-0.4455	480000
0	Telangana	2014	640	530	705	0.486747	615	-1	376	-1	-1	-1	-1	0.99	0.7119	-0.1437	-1.1218	0.8637	550000
1	West Ben	2014	720	630	750	0.338786	485	292	-1	-1	-1	228	-1	1.2772	1.0449	0.7785	-1.2486	0.6721	325000
0	Karnataka	2012	385	515	465	0.525923	415	-1	-1	-1	-1	-1	-1	0.1623	0.9688	1.1074	0.4442	-0.1295	405000
0	Orissa	2011	475	245	485	0.735796	475	-1	-1	-1	-1	-1	-1	-1.3447	-1.3713	-0.4891	0.1798	1.2923	500000
0	Uttar Prad	2014	520	570	620	0.9539	425	-1	530	-1	-1	-1	-1	1.2772	0.0459	0.6248	-1.3753	0.6721	210000
0	Maharash	2013	360	495	280	0.793581	495	-1	-1	-1	-1	-1	-1	-0.1855	-0.1232	-0.4891	-1.2303	-1.3934	400000
0	Tamil Nad	2013	545	505	595	0.356536	555	-1	346	-1	-1	-1	-1	0.2718	0.7119	0.3174	-0.1076	1.4386	360000
0	Orissa	2013	240	340	430	0.143257	295	-1	-1	-1	-1	-1	-1	-1.8825	-2.2851	-0.2974	-0.2344	-0.6692	240000
0	Haryana	2014	440	415	554	0.793581	495	-1	-1	-1	-1	-1	-1	0.335	0.0924	-0.6582	-0.4821	-0.1543	120000
1	Karnataka	2011	355	565	495	0.930371	565	-1	-1	-1	-1	-1	-1	-0.881	0.6568	0.672	1.4724	0.3444	430000
0	Haryana	2011	345	365	445	0.538387	-1	333	-1	-1	-1	-1	-1	-0.4173	-1.3713	0.9042	1.5578	0.3444	120000
1	Maharash	2012	440	440	385	0.14479	435	-1	284	-1	-1	-1	-1	1.1336	-0.2871	-0.1437	0.1459	0.2889	120000
0	Chhattisga	2012	370	410	320	0.488348	405	-1	-1	-1	-1	-1	-1	-0.159	-0.2871	-0.9122	-0.6147	-0.6692	120000

ETL PROCESS:

- ➢ ETL is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system
  - o Full form of ETL is Extract, Transform and Load.
- ➢ The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
  - o Enhances Business Intelligence solutions for decision making.
  - o Allows verification of data transformation, aggregation and calculations rules.
  - o Allows sample data comparison between source and target system.
  - o Helps to improve productivity as it codifies and reuses without additional technical skills.
- ➢ Initially, the raw dataset was arranged as shown in given picture:

ID	Gender	DOB	10percent	10board	12graduat	12percent	12board	CollegeID	CollegeTi	Degree	Specializa	collegeGP	CollegeCi	CollegeCi	CollegeSt	Graduatio
604399	f	22-10-1990	87.8	cbse	2009	84	cbse	6920	1	B.Tech/B.I	instrumer	73.82	6920	1	Delhi	2013
988334	m	15-05-1990	57	cbse	2010	64.5	cbse	6624	2	B.Tech/B.I	computer	65	6624	0	Uttar Prad	2014
301647	m	21-08-1989	77.33	maharash	2007	85.17	amravati c	9084	2	B.Tech/B.I	electronic	61.94	9084	0	Maharash	2011
582313	m	04-05-1991	84.3	cbse	2009	86	cbse	8195	1	B.Tech/B.I	computer	80.4	8195	1	Delhi	2013
339001	f	30-10-1990	82	cbse	2008	75	cbse	4889	2	B.Tech/B.I	biotechnc	64.3	4889	1	Tamil Nad	2012
609356	f	02-12-1989	83.16	icse	2007	77	cbse	10950	1	M.Tech./N	instrumer	99.93	10950	0	Punjab	2013
1081649	f	17-04-1989	72.5	state boar	2007	53.2	state boar	14381	2	B.Tech/B.I	mechanic;	68	14381	1	West Ben	2013
610842	f	11-04-1991	77	state boar	2009	88	state boar	13208	2	B.Tech/B.I	computer	71	13208	1	Telangana	2013
1183070	m	25-11-1992	76.8	state boar	2010	87.7	state boar	5338	2	B.Tech/B.I	informatic	73.15	5338	0	Andhra Pr	2014
794062	f	15-03-1993	57	state boar	2009	73	state boar	8346	2	B.Tech/B.I	computer	70.08	8346	0	Uttar Prad	2014
1088206	m	21-06-1990	77	state boar	2008	75	state boar	13424	2	B.Tech/B.I	electronic	62	13424	0	Maharash	2013
1279958	m	02-07-1992	81.2	state boar	2008	79.9	state boar	64	2	B.Tech/B.I	instrumer	67.67	64	0	Uttar Prad	2013
471413	f	24-12-1991	85	delhi boar	2009	88	all india b	57	2	B.Tech/B.I	informatic	85	57	0	Haryana	2013
1088423	f	09-08-1992	90	state boar	2009	82.1	state boar	2998	2	B.Tech/B.I	computer	85	2998	0	Telangana	2014
1066680	m	12-09-1991	86.4	cbse	2009	86.2	cbse	1906	2	B.Tech/B.I	electronic	81.4	1906	1	West Ben	2014
407672	m	30-09-1990	84.13	0	2008	77	0	3801	2	B.Tech/B.I	informatic	75.2	3801	0	Karnataka	2012
205633	m	25-05-1988	81.7	hse	2005	75.8	chse	5508	2	B.Tech/B.I	computer	78.7	5508	0	Orissa	2011
924541	f	25-02-1992	86	cbse	2010	89	cbse	429	2	B.Tech/B.I	computer	73.9	429	0	Uttar Prad	2014
512353	m	25-10-1991	66.15	state boar	2009	54	state boar	3603	2	B.Tech/B.I	computer	66	3603	0	Maharash	2013
1136577	m	22-05-1991	79.29	icse	2009	68.67	cbse	5298	2	B.Tech/B.I	computer	76	5298	0	Tamil Nad	2013
781327	m	21-11-1987	60	state boar	2006	50	state boar	11013	2	B.Tech/B.I	computer	69.94	11013	0	Orissa	2013
764522	f	14-01-1992	58.4	cbse	2008	64.8	cbse	85	2	MCA	computer	80	85	0	Haryana	2014
203323	m	03-11-1986	61	0	2003	59	0	3716	2	B.Tech/B.I	informatic	63	3716	1	Karnataka	2011
267354	m	05-08-1988	50	board of s	2008	64	state boar	7564	2	B.Tech/B.I	electronic	70	7564	0	Haryana	2011
839626	f	10-04-1991	67.06	state boar	2008	70.67	state boar	13271	2	B.Tech/B.I	computer	55.5	13271	1	Maharash	2012
1169332	f	09-03-1990	67	cbse	2008	61	cbse	16664	2	B.Tech/B.I	computer	75	16664	0	Chhattisga	2012

Steps taken to clean dataset thorough ETL process

Step 1:

➢ Firstly, open blank excel file and go to Data from Tab and select New query and entered data csv file from it.

Step 2:

➢ After entering the query editor, data fields title will promote into header

STEP 4:

➢ Now, from query editor select Home tab and remove column which are not in required for analysis. Column which are removed from dataset are given below:

- College ID
- College Tier
- College City ID
- College City Tier
- English
- Quant
- Domain
- Computer Programming
- Electronics And Semicon
- Computer Science
- Mechanical Engineering
- Electrical Engineering
- Telecom Engineering
- Civil Engineering
- Conscientiousness
- Agreeableness
- Extraversion
- Nueroticism
- Openness To express

Step 5:

- Now to remove null values we first sort the dataset show that null values comes on the top and after that we can delete it through Delete rows from Home tab

`= Table.Sort(#"Removed Columns",{{"10board", Order.Ascending}})`

ID	Gender	DOB	10percentage	10board	12graduation	12perce
4	227582 m	14-10-1987	74		0	2005
5	211840 f	16-08-1989	86		0	2007
6	453953 m	24-09-1985	69.33		0	2004
7	141813 m	31-01-1986	70.5		0	2005
8	337764 m	26-07-1988	68.67		0	2006
9	228148 f	16-09-1989	58		0	2007
10	357969 f	24-07-1991	93		0	2008
11	246543 m	27-06-1989	81.6		0	2006
12	64897 f	13-02-1988	80.64		0	2006
13	353742 m	27-12-1990	76		0	2008
14	390024 m	24-01-1991	73		0	2008
15	121780 f	18-05-1989	92		0	2007
16	336128 m	25-02-1990	87.4		0	2008
17	342743 m	20-12-1991	80.7		0	2008
18	579845 f	05-08-1990	83		0	2008
19	334330 m	20-07-1991	67.2		0	2008
20	226015 m	17-09-1990	87.3		0	2008
21	246938 m	30-05-1988	77		0	2005
22	242078 m	13-02-1989	75.75		0	2007
23	120132 m	22-03-1989	86		0	2006
24	345838 m	29-05-1991	89		0	2008
25	341436 m	25-04-1990	86		0	2007
26	238392 f	24-09-1989	91		0	2007

Query Settings

PROPERTIES
Name
Engineering_graduate_salary
All Properties

APPLIED STEPS
- Source
- Navigation
- Promoted Headers
- Changed Type
- Removed Columns
- X Sorted Rows

`= Table.SelectRows(#"Sorted Rows", each ([10board] <> null and [10board] <> 0 and [10board] <> ""))`

ID	Gender	DOB	10percentage	10board	12graduation	12perce
1	450428 m	08-08-1991	76.4	aisse	2009	
2	349752 m	25-05-1991	80.5	andhra pradesh board ssc	2008	
3	281074 m	11-06-1989	84.67	andhra pradesh state board	2007	
4	354749 m	24-06-1990	82.7	anglo indian	2008	
5	102965 m	02-04-1989	88.5	anglo indian	2006	
6	103861 m	21-07-1988	75.1	anglo indian	2006	
7	480729 m	18-06-1986	83.5	ap state board	2003	
8	456580 m	21-11-1991	89.5	ap state board for secondary education	2008	
9	319678 m	22-07-1991	91	apsche	2008	
10	192591 m	26-03-1990	83	apssc	2007	
11	352532 f	14-06-1990	85.46	aurangabad board	2008	
12	479756 m	10-12-1990	77.6	bharathi matriculation school	2008	
13	604570 m	12-09-1992	62.6	bihar board	2009	
14	345685 f	01-06-1990	75	bihar board	2008	
15	215666 m	21-05-1988	55	bihar board	2005	
16	322718 m	26-01-1990	72	bihar school examination board	2006	
17	389358 m	17-08-1988	59.7	bihar school examination board	2005	
18	814610 m	05-02-1992	76.8	bihar school examination board patna	2008	
19	576054 m	15-11-1988	60	bihar secondary education board,patna	2006	
20	295307 m	06-09-1987	60.14	biharboard	2007	
21	517535 m	01-03-1991	88.83	board of secondary education	2008	
22	248992 m	14-09-1989	67.5	board of high school and intermediate education utta...	2006	
23	243897 f	23-02-1983	70	board of school education harayana	2001	

Query Settings

PROPERTIES
Name
Engineering_graduate_salary
All Properties

APPLIED STEPS
- Source
- Navigation
- Promoted Headers
- Changed Type
- Removed Columns
- Sorted Rows
- X Filtered Rows

Step 6:

- Finally, sort the dataset through filter of 10th percentage data fields

	ID	Gender	DOB	10percentage	10board	12graduation	12perce
1	619705	m	20-10-1991	97.76	state board	2009	
2	239868	f	28-05-1989	97.12	karnataka secondary education examination board	2007	
3	1258784	m	20-01-1993	97	state board	2010	
4	1203568	f	20-04-1993	96.8	cbse	2010	
5	555599	f	20-07-1991	96.48	state board	2009	
6	94788	f	08-04-1988	96	stjosephs girls higher sec school,dindigul	2005	
7	924753	m	01-09-1992	96	icse	2010	
8	1171203	m	28-08-1992	96	icse	2010	
9	1230582	m	12-10-1992	95.8	state board	2010	
10	716400	f	28-11-1991	95.6	cbse	2009	
11	877241	m	10-05-1992	95.54	state board	2010	
12	607667	f	31-03-1991	95.52	karnataka secondary education examination board	2009	
13	192948	f	26-04-1989	95.4	cbse	2007	
14	606571	m	06-05-1991	95.2	cbse	2009	
15	811167	m	10-02-1992	95.2	cbse	2009	
16	901358	f	03-03-1993	95.2	cbse	2010	
17	361671	m	11-10-1990	95.04	kseeb	2008	
18	1106060	f	04-04-1992	95	cbse	2010	
19	1229814	f	07-01-1992	95	state board	2009	
20	265303	f	06-10-1989	94.8	state board	2007	
21	963241	m	22-08-1993	94.8	state board	2010	
22	595845	f	10-12-1991	94.8	icse	2010	
23	642309	m	18-03-1991	94.8	cbse	2008	
24							

= Table.Sort(#"Filtered Rows",{{"10percentage", Order.Descending}})

Query Settings

PROPERTIES
Name
Engineering_graduate_salary
All Properties

APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Removed Columns
Sorted Rows
Filtered Rows
X Sorted Rows1

Step 7:

- Remove duplicates from the data using duplicate option rom home tab

	ID	Gender	DOB	10percentage	10board	12graduation	12perce
1	619705	m	20-10-1991	97.76	state board	2009	
2	239868	f	28-05-1989	97.12	karnataka secondary education examination board	2007	
3	1258784	m	20-01-1993	97	state board	2010	
4	1203568	f	20-04-1993	96.8	cbse	2010	
5	555599	f	20-07-1991	96.48	state board	2009	
6	94788	f	08-04-1988	96	stjosephs girls higher sec school,dindigul	2005	
7	924753	m	01-09-1992	96	icse	2010	
8	1171203	m	28-08-1992	96	icse	2010	
9	1230582	m	12-10-1992	95.8	state board	2010	
10	716400	f	28-11-1991	95.6	cbse	2009	
11	877241	m	10-05-1992	95.54	state board	2010	
12	607667	f	31-03-1991	95.52	karnataka secondary education examination board	2009	
13	192948	f	26-04-1989	95.4	cbse	2007	
14	606571	m	06-05-1991	95.2	cbse	2009	
15	811167	m	10-02-1992	95.2	cbse	2009	
16	901358	f	03-03-1993	95.2	cbse	2010	
17	361671	m	11-10-1990	95.04	kseeb	2008	
18	1106060	f	04-04-1992	95	cbse	2010	
19	1229814	f	07-01-1992	95	state board	2009	
20	265303	f	06-10-1989	94.8	state board	2007	
21	963241	m	22-08-1993	94.8	state board	2010	
22	595845	f	10-12-1991	94.8	icse	2010	
23	642309	m	18-03-1991	94.8	cbse	2008	
24							

= Table.Distinct(#"Sorted Rows1", {"ID"})

Query Settings

PROPERTIES
Name
Engineering_graduate_salary
All Properties

APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Removed Columns
Sorted Rows
Filtered Rows
Sorted Rows1
X Removed Duplicates

Finally, after cleaning the data, the final dataset sample is shown below

ID	Gender	DOB	10percentage	10board	12graduation	12percentage	12board
619705	m	20-10-1991	97.76	state board	2009	96.33	state board
239868	f	28-05-1989	97.12	karnataka secondary education examination board	2007	90.33	department of pre-university education(governm
1258784	m	20-01-1993	97	state board	2010	95.08	state board
1203568	f	20-04-1993	96.8	cbse	2010	92.3	cbse
555599	f	20-07-1991	96.48	state board	2009	82	state board
924753	m	01-09-1992	96	icse	2010	79.5	state board
94788	f	08-04-1988	96	stjosephs girls higher sec school,dindigul	2005	92	srv girls higher sec school,rasipuram
1171203	m	28-08-1992	96	icse	2010	87	cbse
1230582	m	12-10-1992	95.8	state board	2010	87.16	state board
716400	f	28-11-1991	95.6	cbse	2009	94.4	cbse
877241	m	10-05-1992	95.54	state board	2010	90	state board
607667	f	31-03-1991	95.52	karnataka secondary education examination board	2009	94.33	department of pre-university education, bangalor
192948	f	26-04-1989	95.4	cbse	2007	93.2	cbse
606571	m	06-05-1991	95.2	cbse	2009	91.4	cbse
811167	m	10-02-1992	95.2	cbse	2009	78.3	cbse
901358	f	03-03-1993	95.2	cbse	2010	74.8	cbse
361671	m	11-10-1990	95.04	kseeb	2008	87	karnataka pu board
1229814	f	07-01-1992	95	state board	2009	85	state board
1106060	f	04-04-1992	95	cbse	2010	89	cbse
265303	f	06-10-1989	94.8	state board	2007	84.09	state board
595845	f	10-12-1991	94.8	icse	2010	93	icse
642309	m	18-03-1991	94.8	cbse	2008	81.2	cbse
963241	m	22-08-1993	94.8	state board	2010	97.4	state board
47476	f	10-06-1988	94.7	cbse	2006	94.7	state board
900799	m	15-03-1992	94.6	state board	2009	91.2	state board
190981	m	18-11-1989	94.6	icse	2006	83	cbse

Analysis on dataset 1.

HIGHEST SALARY ON THE BASIS OF SPECIALIZATION

Introduction

➢ By performing this analysis, we will get the name of Specification in which the engineering candidate get highest salary after graduating.

Description

➢ The analysis is based on Specialization and salary of the candidate.
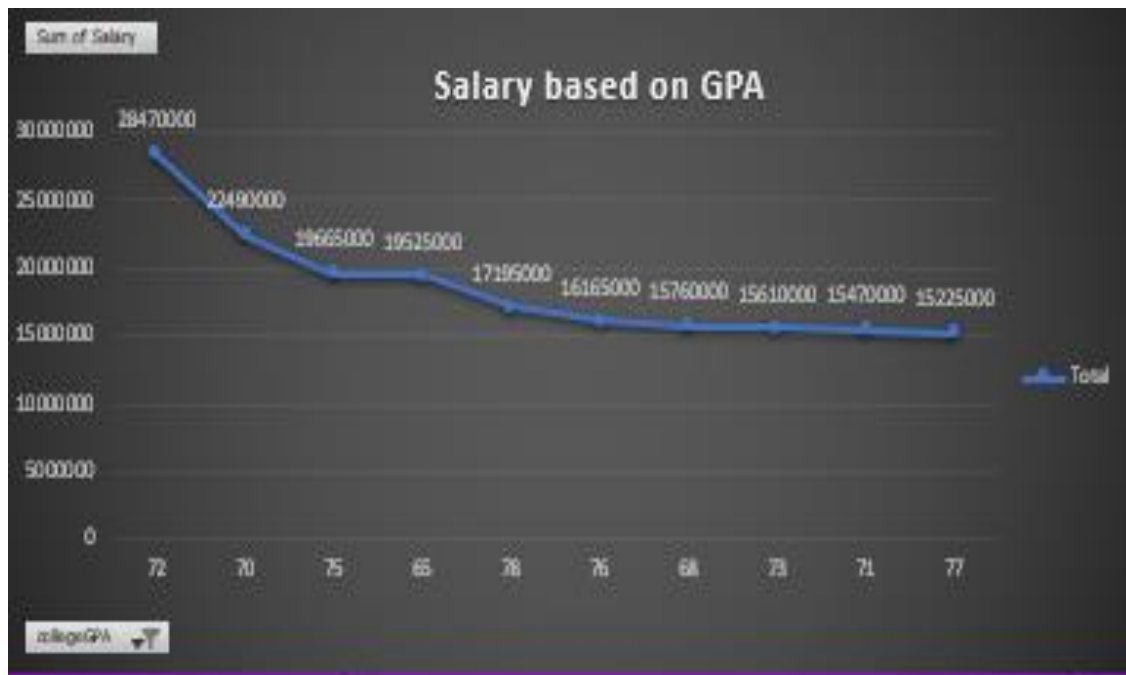
Specific requirements, functions and formulas

➢ Pivot table is used for the analysis.
➢ Sum function is used in pivot table for the sum of the salary in the pivot table.

Analysis results

Row Labels	Sum of Salary
electronics and communication engineering	193070000
computer engineering	159750000
computer science & engineering	155215000
information technology	154040000
computer application	55935000
Grand Total	718010000

Visualization

2. SALARY BASED ON GPA

INTRODUCTION
- ➢ By performing this analysis, we will get to know about the candidates who got the highest salary after graduating how much they scored in Logical.

Description
- ➢ The analysis based on the GPA, salary of the candidates, functions and formulas.

Specific requirements, functions and formulas
- ➢ Pivot table is used for the analysis.
- ➢ Sum function is used in pivot table for the sum of the salary in the pivot table

Analysis results

Row Labels	Sum of Salary
72	28470000
70	22490000
75	19665000
65	19525000
78	17195000
76	16165000
68	15760000
73	15610000
71	15470000
77	15225000
Grand Total	185575000

Visualization

3.SALARY BASED ON LOGICAL

Introduction:

  ➢ By performing this analysis, we will get to know about the candidates who got the highest salary after graduating how much they scored in Logical.

Description

  ➢ The analysis based on the Logical result, salary of the candidate, functions and formulas.
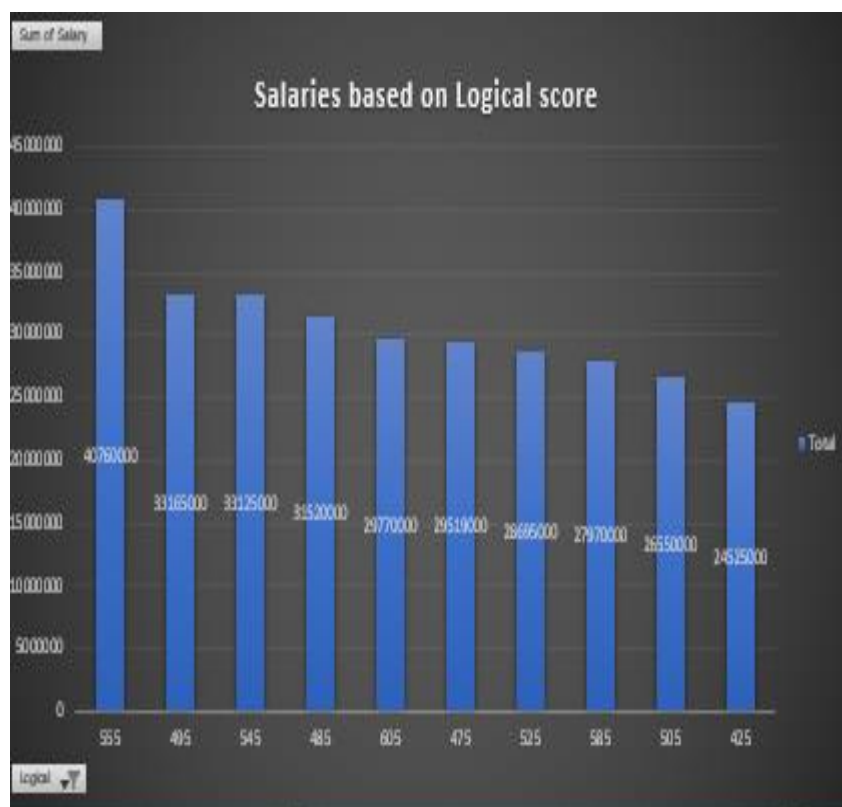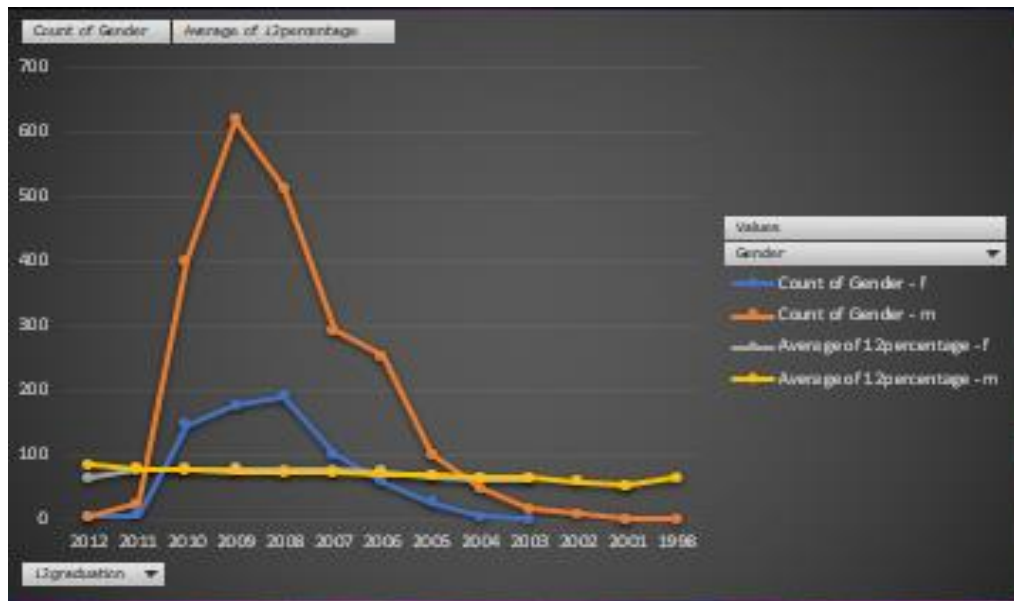
Specific requirements, functions and formulas

  ➢ Pivot table is used for the analysis.
  ➢ Sum function is used in pivot table for the sum of the salary in the pivot table

Analysis results

Row Labels	Sum of Salary
555	40760000
495	33165000
545	33125000
485	31520000
605	29770000
475	29519000
525	28695000
585	27970000
505	26550000
425	24525000
Grand Total	**305599000**

Visualization

4.EXPLOATION ON 12TH BOARD

INTRODUCTION

 ➤ By performing this analysis, we will get to know year wise average result of 12th board as per male and female.

 Description

 ➤ The analysis based on the gender, 12th Graduation ,12th Percentage of the candidates, functions and formulas.

 Specific requirements, functions and formulas

 ➤ Pivot table is used for the analysis.
 ➤ Count and Average function is used in pivot table for the count of number of male and female and to find the average of 12th Percentage in the pivot table

 Analysis results

Row Labels	Count of Gender f	m	Average of 12percentage f	m	Total Count of Gender	Total Average of 12percentage
2012	4	5	65.375	84.45	9	75.97222222
2011	7	24	76.95714286	77.92875	31	77.70935484
2010	147	401	79.60326531	76.56476309	548	77.37983577
2009	176	617	78.85818182	75.50016207	793	76.24544767
2008	190	512	77.05484211	72.74027344	702	73.90803419
2007	101	293	76.03227723	72.13412969	394	73.13340102
2006	58	253	74.855	70.96011858	311	71.68649518
2005	26	101	67.17692308	68.29752475	127	68.06811024
2004	5	49	63.23	64.20306122	54	64.11296296
2003	1	16	62.16	65.56875	17	65.36823529
2002		9		59.69	9	59.69
2001	1	1	50	52	2	51
1998		1		65.25	1	65.25
Grand Total	716	2282	77.11885475	73.46950044	2998	74.34106071

Visualization

5.EXPLOATION ON 10TH BOARD

INTRODUCTION
- ➤ By performing this analysis, we will get top 10 state which scored the highest result in 10th board.

Description
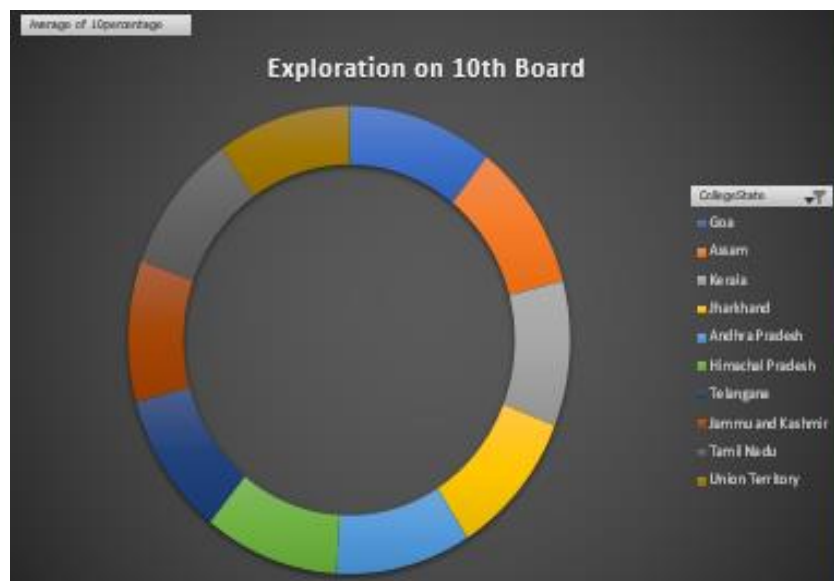- ➤ The analysis based on the Collage State, 10 Percentage of the candidates, functions and formulas.

Specific requirements, functions and formulas
- ➤ Pivot table is used for the analysis.
- ➤ Average function is used in pivot table to find the average of 10th Percentage in the pivot table

Analysis results

Row Labels	Average of 10percentage
Goa	89
Assam	84.994
Kerala	83.56
Jharkhand	83.4615
Andhra Pradesh	82.4972093
Himachal Pradesh	82.38461538
Telangana	82.1597992
Jammu and Kashmir	81.24
Tamil Nadu	81.18833333
Union Territory	80.96
Grand Total	**81.9942876**

Visualization

5.STATEWISE ANALYSIS ON DEGREE

INTRODUCTION

➢ By performing this analysis, we will get the degree name which is highest taken by the students in each state.

Description

➢ The analysis based on the State, Gender, Degree functions and formulas.

Specific requirements, functions and formulas

➢ Pivot table is used for the analysis.
➢ Count function is used in pivot table to find the number of students in each degree as per the state in the pivot table

Analysis results

Count of Gender	Column Labels				
Row Labels	B.Tech/B.E.	M.Sc. (Tech.)	M.Tech./M.E.	MCA	Grand Total
Uttar Pradesh	618		6	37	661
Karnataka	217		8	18	243
Telangana	219		2	12	233
Tamil Nadu	209		4	13	226
Maharashtra	143		4	11	158
Andhra Pradesh	145		2	8	155
Madhya Pradesh	127		1	7	135
West Bengal	128		2	4	134
Punjab	114		4	14	132
Haryana	126		1	3	130
Delhi	107			12	119
Rajasthan	113		1	2	116
Orissa	103			7	110
Uttarakhand	66			10	76
Kerala	22				22
Jharkhand	14			5	19
Gujarat	15	1	2		18
Chhattisgarh	17				17
Grand Total	2503	1	37	163	2704

## Visualization
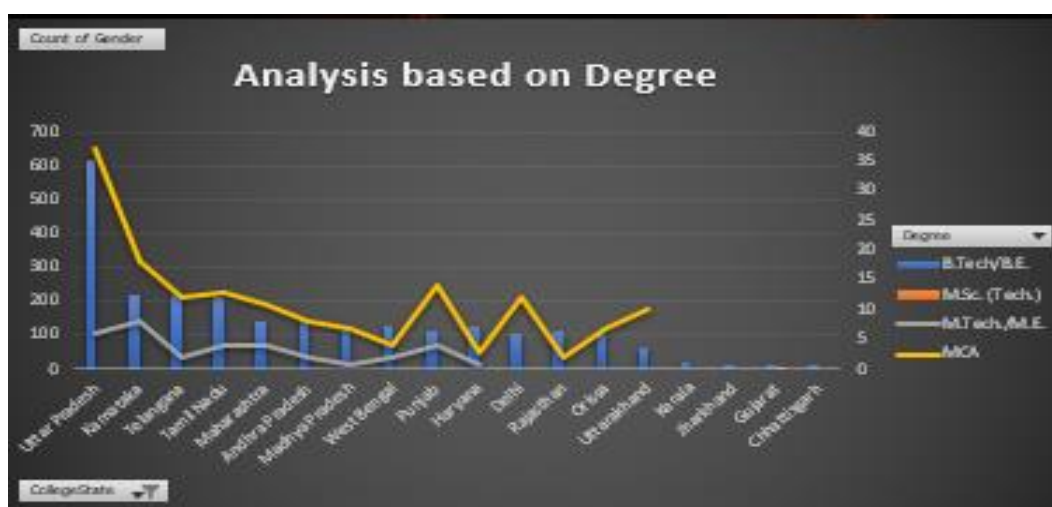
List of Analysis with results

1.HIGHEST SALARY ON THE BASIC OF SPECIALIZATION

- ➢ The highest Salary is recorded in electronics and communication engineering with total of Rs 193070000.
- ➢ After that the second highest is recorded in Computer engineering with total of Rs15975000 .


2.SALARY BASED ON GPA

- ➢ The highest Salary after graduating was recorded to candidate who got only 72 GPA in collage.

- ➢ After that the second highest salary was recorded to candidate who got 70 GPA in collage.
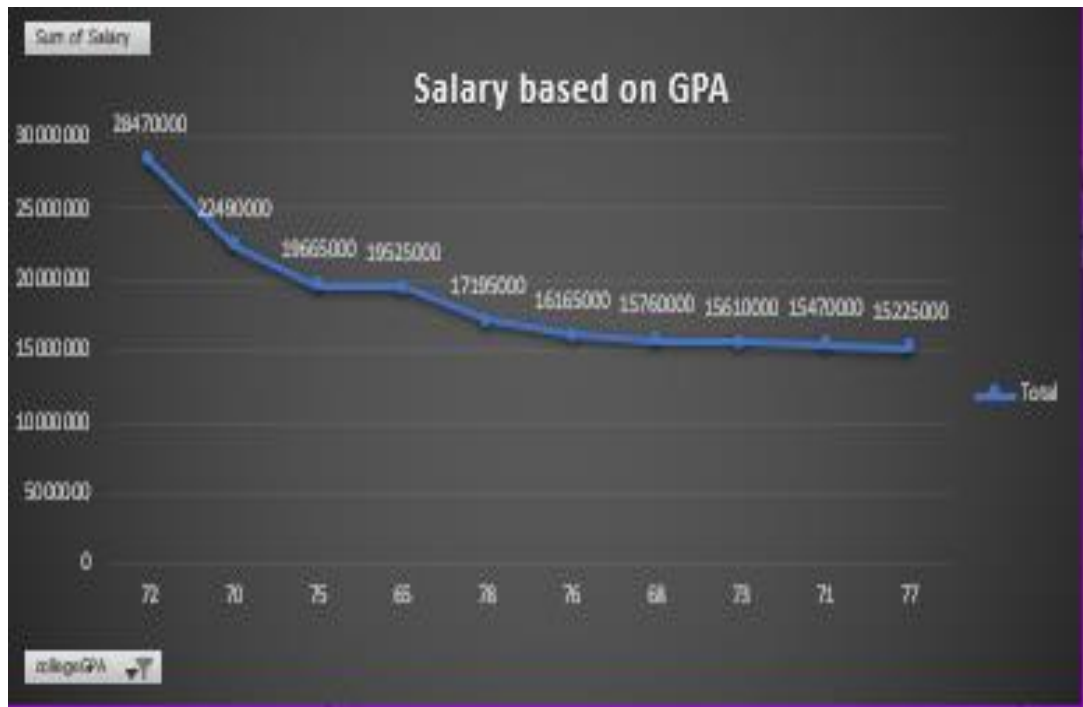
Salary based on GPA

3.SALARY BASED ON LOGICAL

➢ The highest Salary was recorded to candidate who got 555 marks in Logical Test getting total salary of Rs40760000

➢ After that the second highest salary was recorded to candidate who got 495 marks in Logical Test getting total salary of RS33165000.
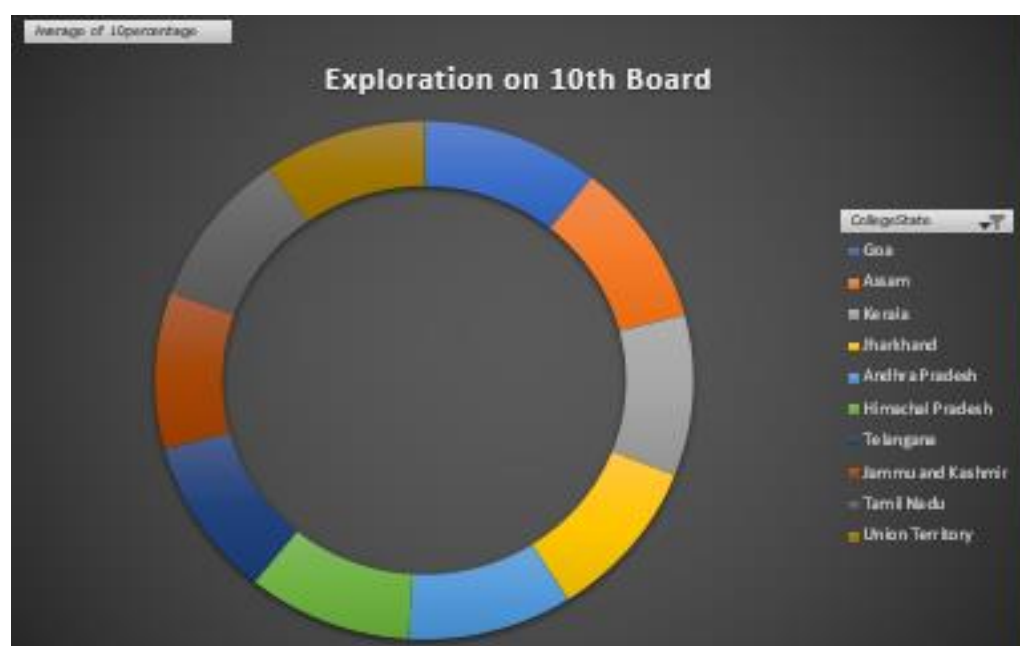

Salaries based on Logical score
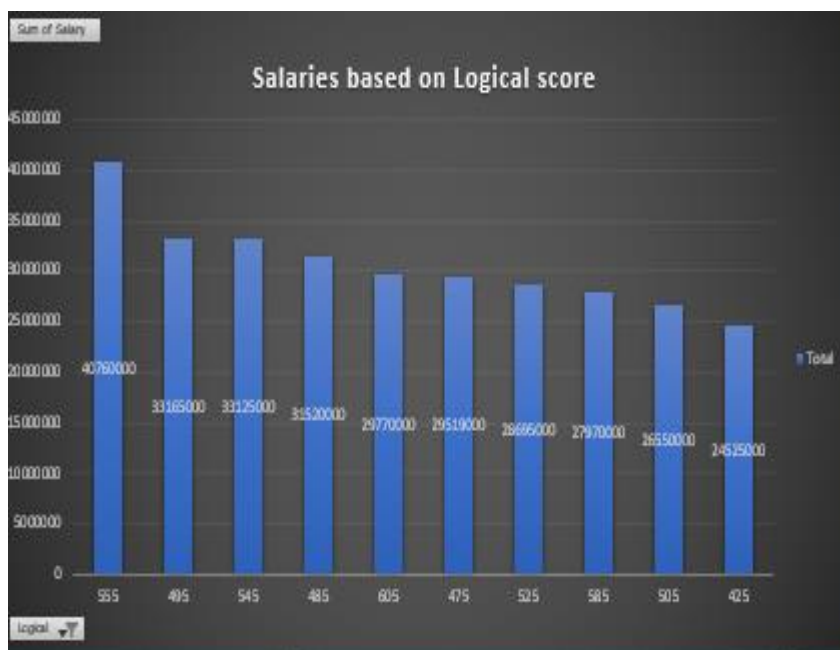
4.EXPLOATION ON 12TH BOARD

➢ The highest percentage of class 12th was recorded in the year 2011 which is 77.70. Whereas lowest was recorded in the year 2001 which is 51.
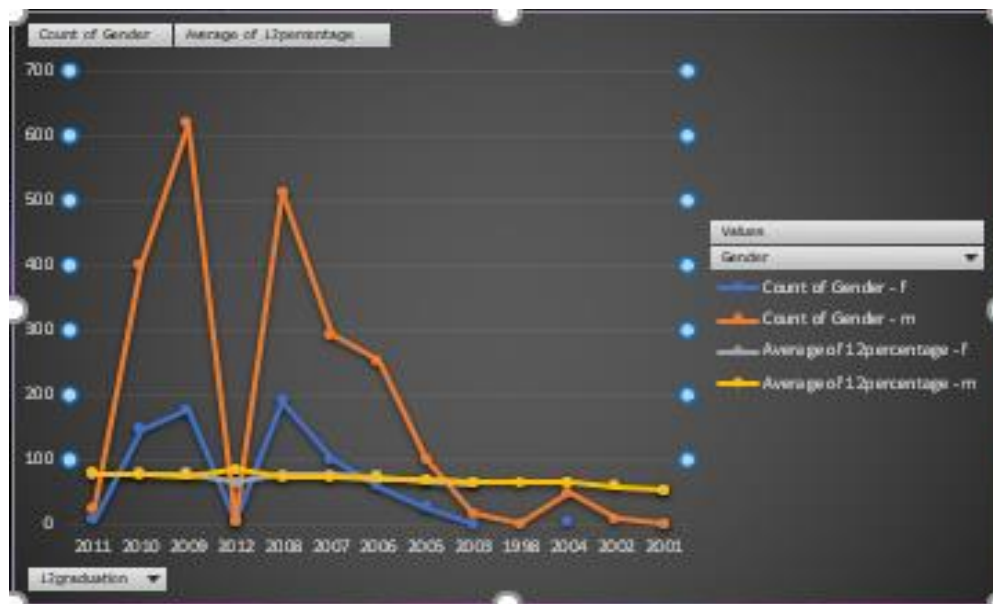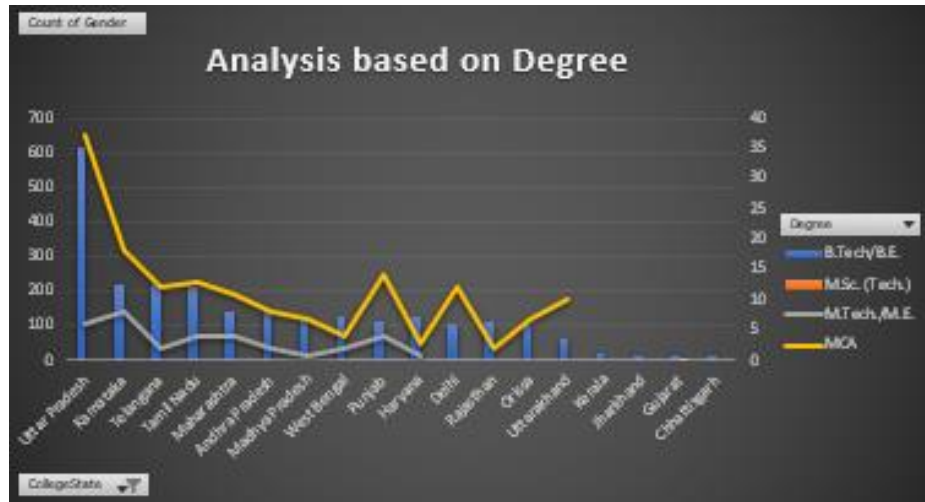

5.EXPLORATION ON 10TH BOARD

➢ The highest percentage of class 10th was recorded in the state GOA which is 89. Whereas lowest was recorded in Union territory which is 80.

6. STATEWISE ANALYSIS ON DEGREE

The Highest number of students taking the degree of B.tech/BE which is 2503 and the state from where higher number of students are graduating is ARUNACHAL PRADESH which is 661.

FINAL DASHBOARD:

BIBLIOGRAPHY:

Dataset source:

https://www.kaggle.com/manishkc06/engineering-graduate-salary-prediction

Dashboard Background image:
https://tse1.mm.bing.net/th?id=OIP.sM3b14VYeLQBxLouHy9GTAHaEo&pid=Api&P=0&w=248&h=156

Information about Data Management:

https://www.blue-pencil.ca/what-is-data-management-and-why-it-is-important/