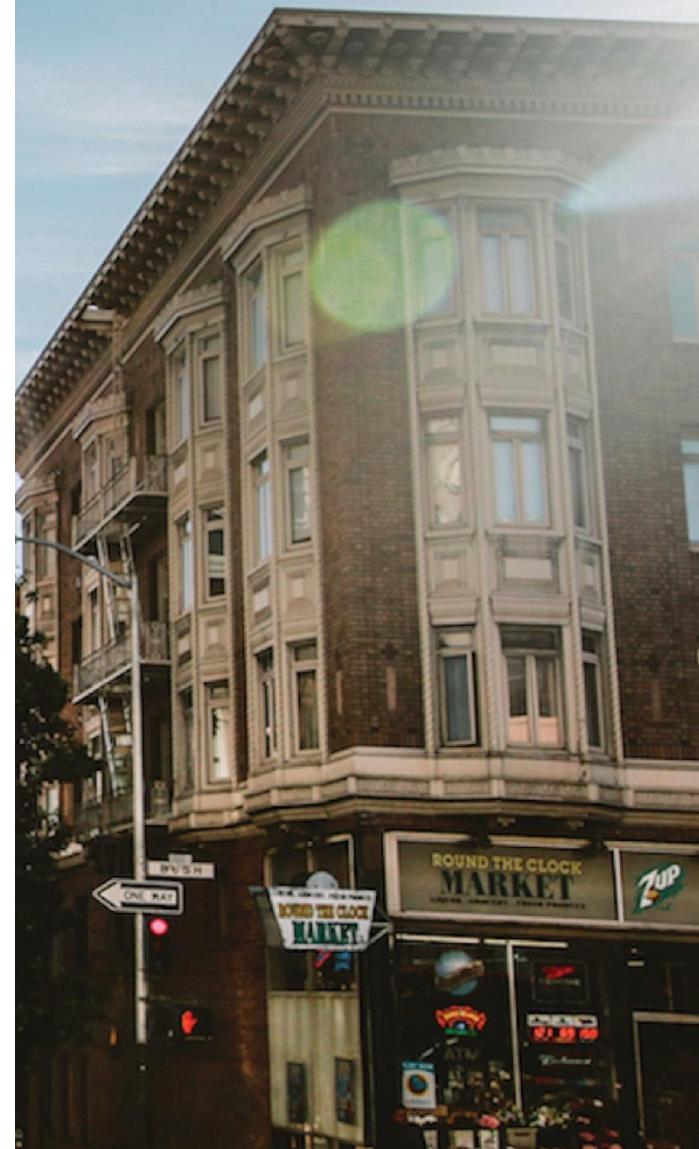


Data Analysis Portfolio

AUGUST 25

By Aastha Kumar



Professional Background

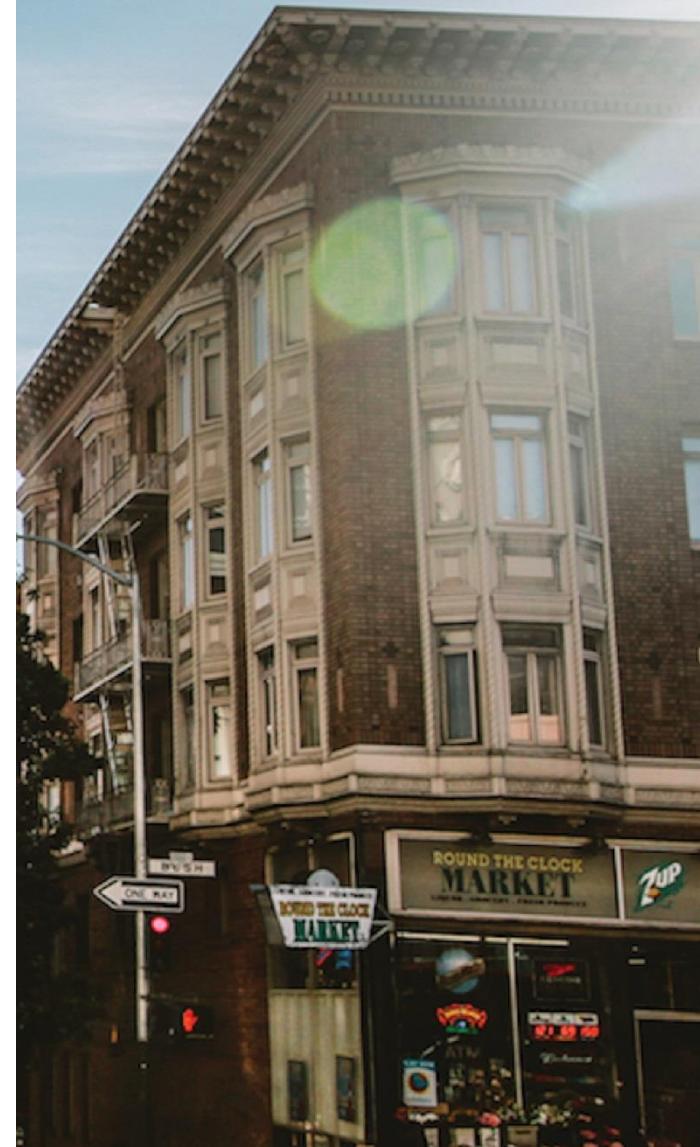
I am a student at Vellore Institute of Technology currently pursuing a B.Tech in Computer Science and Engineering. I enjoy solving technical problems, researching and developing new technologies, designing software applications for different platforms. I enjoy meeting people and working with them in a team environment. I am a quick learner with a fun, outgoing personality. In addition, I excel in my ability to work under pressure and handle stressful situations very well. Currently looking for like-minded individuals and companies to connect with in order to gain and share knowledge in similar areas of interest.



Instagram User Analytics

JULY 27

By Aastha Kumar

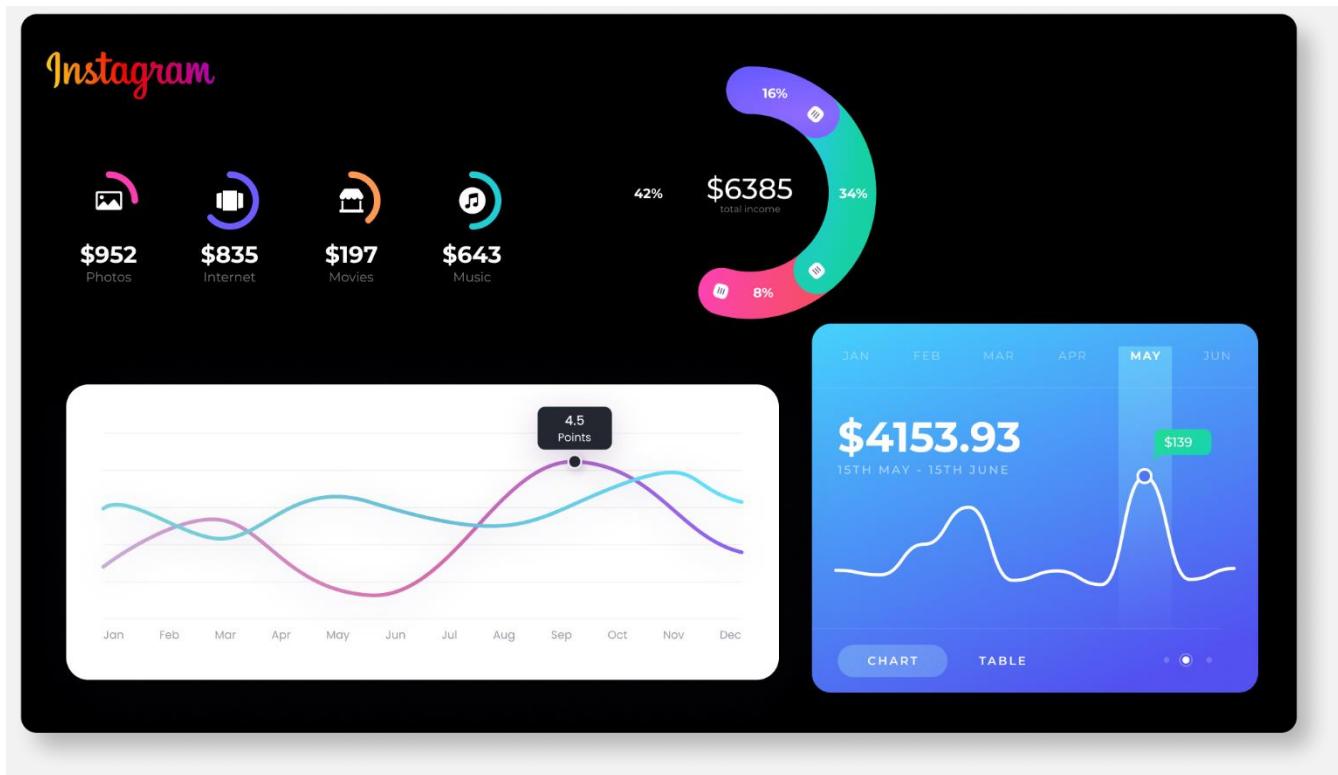


Description

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams.

These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.

You are working with the product team of Instagram and the product manager has asked you to provide insights on the questions asked by the management team.



Approach

First, I started off by cloning the dataset into my device using MySql. I then ran a series of commands to filter out data from the database according to the criteria demanded by the user.

Tech-Stack Used

- MySql
- MySql Work Bench
- Google Docs

Insights

A) Marketing: The marketing team wants to launch some campaigns, and they need your help with the following

1. Rewarding Most Loyal Users: People who have been using the platform for the longest time.

What I Did: Find the 5 oldest users of the Instagram from the database provided

```
mysql> SELECT * FROM users ORDER BY created_at ASC LIMIT 5;
+----+-----+-----+
| id | username          | created_at        |
+----+-----+-----+
| 80 | Darby_Herzog     | 2016-05-06 00:14:21 |
| 67 | Emilio_Bernier52 | 2016-05-06 13:04:30 |
| 63 | Elenor88          | 2016-05-08 01:30:41 |
| 95 | Nicole71          | 2016-05-09 17:30:22 |
| 38 | Jordyn.Jacobson2 | 2016-05-14 07:56:26 |
+----+-----+-----+
5 rows in set (0.00 sec)
```

-
- 2. Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo.

What I Did: Find the users who have never posted a single photo on Instagram

```
mysql> SELECT * FROM users WHERE users.id NOT IN (SELECT user_id FROM photos);
+----+-----+-----+
| id | username          | created_at        |
+----+-----+-----+
|  5 | Aniya_Hackett     | 2016-12-07 01:04:39 |
|  7 | Kasandra_Homenick | 2016-12-12 06:50:08 |
| 14 | Jaclyn81          | 2017-02-06 23:29:16 |
| 21 | Rocio33           | 2017-01-23 11:51:15 |
| 24 | Maxwell.Halvorson | 2017-04-18 02:32:44 |
| 25 | Tierra.Trantow    | 2016-10-03 12:49:21 |
| 34 | Pearl7             | 2016-07-08 21:42:01 |
| 36 | Ollie_Ledner37    | 2016-08-04 15:42:20 |
| 41 | Mckenna17          | 2016-07-17 17:25:45 |
| 45 | David.Osinski47    | 2017-02-05 21:23:37 |
| 49 | Morgan.Kassulke   | 2016-10-30 12:42:31 |
| 53 | Linnea59           | 2017-02-07 07:49:34 |
| 54 | Duane60             | 2016-12-21 04:43:38 |
| 57 | Julien_Schmidt    | 2017-02-02 23:12:48 |
| 66 | Mike.Auer39         | 2016-07-01 17:36:15 |
| 68 | Franco_Keebler64  | 2016-11-13 20:09:27 |
| 71 | Nia_Haag            | 2016-05-14 15:38:50 |
| 74 | Hulda.Macejkovic  | 2017-01-25 17:17:28 |
| 75 | Leslie67            | 2016-09-21 05:14:01 |
| 76 | Janelle.Nikolaus81 | 2016-07-21 09:26:09 |
| 80 | Darby_Herzog        | 2016-05-06 00:14:21 |
| 81 | Esther.Zulauf61    | 2017-01-14 17:02:34 |
| 83 | Bartholome.Bernhard | 2016-11-06 02:31:23 |
| 89 | Jessyca_West        | 2016-09-14 23:47:05 |
| 90 | Esmeralda.Mraz57   | 2017-03-03 11:52:27 |
| 91 | Bethany20           | 2016-06-03 23:31:53 |
+----+-----+-----+
26 rows in set (0.02 sec)
```

- 3. Declaring Contest Winner: The team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.

What I Did: Identify the winner of the contest and provide their details to the team

- 4. Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.

What I Did: Identify and suggest the top 5 most commonly used hashtags on the platform

```
mysql> SELECT TAG_ID FROM PHOTO_TAGS GROUP BY TAG_ID ORDER BY COUNT(TAG_ID) DESC LIMIT 5;
+-----+
| TAG_ID |
+-----+
|    21 |
|    20 |
|    17 |
|    13 |
|    18 |
+-----+
5 rows in set (0.00 sec)

mysql> SELECT * FROM TAGS WHERE ID IN (21,20,17,13,18);
+----+-----+-----+
| id | tag_name | created_at      |
+----+-----+-----+
| 13 | fun      | 2022-07-26 13:08:44 |
| 17 | party    | 2022-07-26 13:08:44 |
| 18 | concert   | 2022-07-26 13:08:44 |
| 20 | beach    | 2022-07-26 13:08:44 |
| 21 | smile    | 2022-07-26 13:08:44 |
+----+-----+-----+
5 rows in set (0.00 sec)
```

5. Launch AD Campaign: The team wants to know, which day would be the best day to launch ADs.

What I Did: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

```
mysql> SELECT DAYOFWEEK(created_at) FROM users GROUP BY created_at ORDER BY COUNT(created_at) DESC LIMIT 1;
+-----+
| DAYOFWEEK(created_at) |
+-----+
|            5          |
+-----+
1 row in set (0.00 sec)
```

5th day of the week i.e Thursaday

- B) Investor Metrics: Our investors want to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

-
1. User Engagement: Are users still as active and post on Instagram or they are making fewer posts

What I Did: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

```
mysql> SELECT COUNT(ID) FROM USERS;
+-----+
| COUNT(ID) |
+-----+
|      100 |
+-----+
1 row in set (0.00 sec)

mysql> SELECT COUNT(IMAGE_URL) FROM PHOTOS;
+-----+
| COUNT(IMAGE_URL) |
+-----+
|        257 |
+-----+
1 row in set (0.00 sec)
```

2. Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

What I Did: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

Resources



[SQL Installation Resources](#)



[Dataset for SQL Database](#)

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

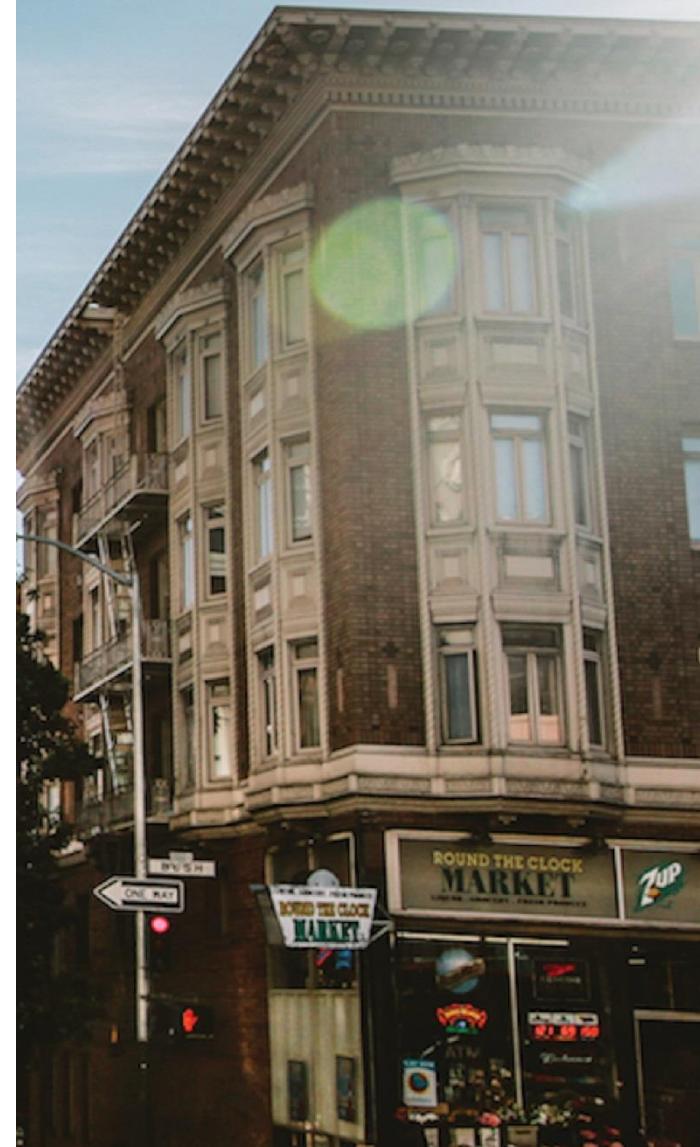
The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.

Operation Analytics and Investigating

AUGUST 1

By Aastha Kumar



Description

Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc and help them derive insights out of the data they collect.

Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.

Investigating metric spike is also an important part of operation analytics as being a Data Analyst you must be able to understand or make other teams understand questions like- Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that its very important to investigate metric spike.



Approach

First, I started off by cloning the dataset into my device using MySql. I then ran a series of commands to filter out data from the database according to the criteria demanded by the user.

Tech-Stack Used

- MySql
- MySql Work Bench
- Google Docs

Insights

Case Study 1 (Job Data)

A. Number of jobs reviewed: Amount of jobs reviewed over time.

What I Did: Calculate the number of jobs reviewed per hour per day for November 2020?

B. Throughput: It is the no. of events happening per second.

What I Did: Let's say the above metric is called throughput. Calculate 7 day rolling average of throughput? For throughput, do you prefer daily metric or 7-day rolling and why?

C. Percentage share of each language: Share of each language for different contents.

What I Did: Calculate the percentage share of each language in the last 30 days?

D. Duplicate rows: Rows that have the same value present in them.

What I Did: Let's say you see some duplicate rows in the data. How will you display duplicates from the table?

Case Study 2 (Investigating metric spike)

- A. User Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service.
What I Did: Calculate the weekly user engagement?
- B. User Growth: Amount of users growing over time for a product.
What I Did: Calculate the user growth for product?
- C. Weekly Retention: Users getting retained weekly after signing-up for a product.
What I Did: Calculate the weekly retention of users-sign up cohort?
- D. Weekly Engagement: To measure the activeness of a user. Measuring if the user finds quality in a product/service weekly.
What I Did: Calculate the weekly engagement per device?
- E. Email Engagement: Users engaging with the email service.
What I Did: Calculate the email engagement metrics?

Resources



[Case Study-1 Dataset](#)



[Case Study-2 Dataset](#)

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

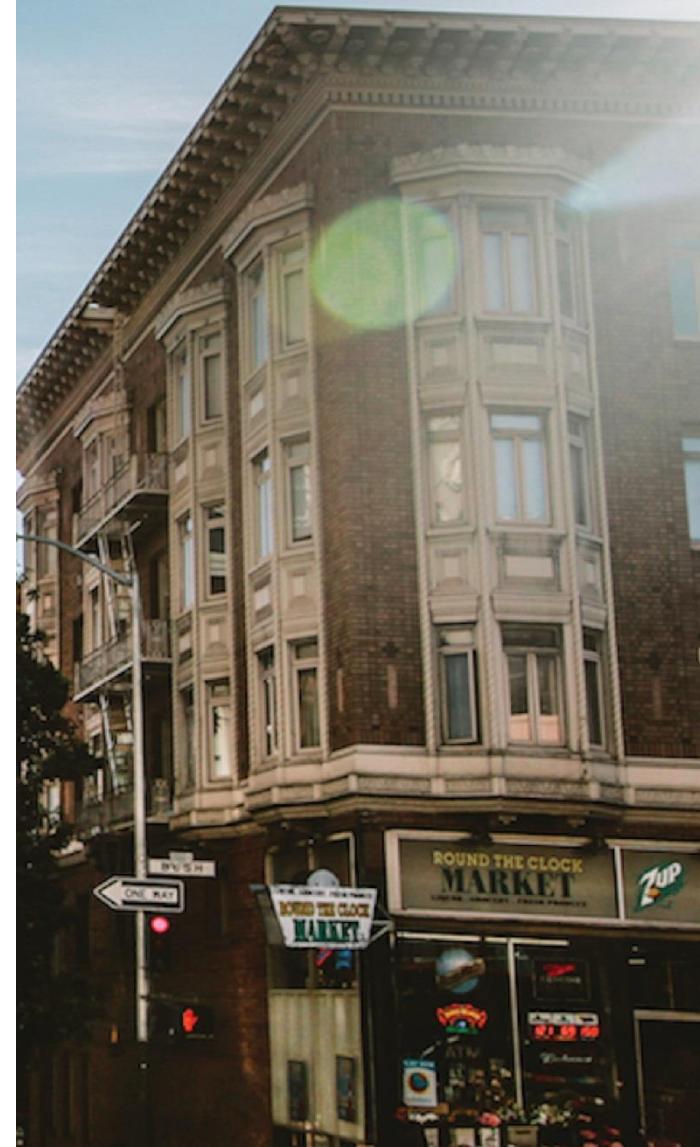
The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.

Hiring Process Analytics

JULY 28

By Aastha Kumar



Description

Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process. Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyze before hiring freshers or any other individual. Thus, making an opportunity for a Data Analyst job here too!



Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

Tech-Stack Used

- Google Sheets

Insights

A. Hiring: Process of intaking of people into an organization for different kinds of positions.

My task: How many males and females are Hired?

Using the filter option we can sort the males and females separately and then use COUNT() to get the individual count of each

The figure consists of three side-by-side screenshots of a Google Sheets interface. The leftmost screenshot shows the 'Filter' dialog open over a table with columns 'Status', 'event_name', and 'Dept'. The middle screenshot shows the 'Filter' dialog again, but with the 'Female' checkbox selected. The rightmost screenshot shows the final result: a table where all 'Male' entries have been removed, leaving only 'Female' entries. The columns are labeled 'Status', 'event_name', and 'Dept'.

Status	event_name	Dept
0 Hi	Sort A to Z	Ser
8 Hi	Sort Z to A	Ser
8 Re	Sort by Color	Ser
8 Re	Sort by Color	Op
2 Hi	Sheet View	Op
4 Hi	Clear Filter From "event_name"	Sal
7 Re	Filter by Color	Sal
7 Re	Text Filters	Ser
9 Hi	Search	Ser
1 Re	(Select All)	Ser
0 Re	<input checked="" type="checkbox"/>	Ser
8 Hi	-	Ser
0 Hi	<input checked="" type="checkbox"/>	Ser
1 Hi	Don't want to say	Fin
8 Hi	<input checked="" type="checkbox"/>	Ser
8 Hi	Female	Ser
8 Hi	<input checked="" type="checkbox"/>	Ser
7 Hi	Male	Ser
1 Re	OK	Op
3 Hi	Cancel	Op
4 Re		Op
E Re		On

Status	event_name	Dept
0 Hi	Sort A to Z	Serv
8 Hi	Sort Z to A	Oper
8 Re	Sort by Color	Serv
2 Hi	Sheet View	Serv
4 Hi	Clear Filter From "event_name"	Service
7 Re	Filter by Color	Service
7 Re	Text Filters	Operal
9 Hi	Search	Operal
1 Re	(Select All)	Serv
0 Re	<input checked="" type="checkbox"/>	Serv
8 Hi	-	Serv
0 Hi	<input checked="" type="checkbox"/>	Purc
1 Hi	Don't want to say	Final
8 Hi	<input checked="" type="checkbox"/>	Final
8 Hi	Female	Final
8 Hi	<input checked="" type="checkbox"/>	Final
7 Hi	Male	Final
1 Re	OK	Oper
3 Hi	Cancel	Oper
4 Re		Final
E Re		Final

Status	event_name	Dept
0 Hi	Sort A to Z	Female
8 Hi	Sort Z to A	Female
8 Re	Sort by Color	Female
2 Hi	Sheet View	Female
4 Hi	Clear Filter From "event_name"	Female
7 Re	Filter by Color	Female
7 Re	Text Filters	Female
9 Hi	Search	Female
1 Re	(Select All)	Sales
0 Re	<input checked="" type="checkbox"/>	Sales
8 Hi	-	Sales
0 Hi	<input checked="" type="checkbox"/>	Service
1 Hi	Don't want to say	Service
8 Hi	<input checked="" type="checkbox"/>	Purcha
8 Hi	Female	Financ
8 Hi	<input checked="" type="checkbox"/>	Financ
7 Hi	Male	Financ
1 Re	OK	Operal
3 Hi	Cancel	Operal
4 Re		Final
E Re		Final

Thus, no. of females = 2675 and no. of males = 4085

B. Average Salary: Adding all the salaries for a select group of employees and then dividing the sum by the number of employees in the group.

My task: What is the average salary offered in this company?

Statistics - Excel

Aastha Kumar AK

Tell me what you want to do

File Home Insert Page Layout Formulas Data Review View Help

Font Alignment Number Conditional Formatting Format as Table Cell Styles Insert Delete Format Cells Editing

G7170 : fx =AVERAGE(G2:G7169)

7152	641060	20-08-2014 17:49	Rejected	Male	Operations Department	i6	67779
7153	786222	20-08-2014 17:51	Hired	Male	Operations Department	i6	47286
7154	964206	25-08-2014 09:23	Hired	Male	Service Department	i5	7849
7155	586934	29-08-2014 19:00	Hired	Don't want to say	Service Department	c9	63159
7156	321440	17-08-2014 12:17	Hired	Female	Production Department	c9	89565
7157	975415	17-08-2014 07:48	Hired	Male	Service Department	c5	86162
7158	683982	17-08-2014 07:49	Hired	Don't want to say	Service Department	c5	58900
7159	665614	21-08-2014 07:30	Rejected	Male	Sales Department	c5	71449
7160	935497	27-08-2014 17:36	Rejected	Male	Service Department	c5	67196
7161	106032	22-08-2014 15:50	Hired	Male	Service Department	i7	16756
7162	166185	27-08-2014 05:01	Hired	Female	Service Department	i7	30952
7163	736189	28-08-2014 17:29	Hired	Male	Service Department	c9	64150
7164	614594	28-08-2014 17:30	Hired	Male	Service Department	c9	40152
7165	493131	28-08-2014 17:32	Hired	Male	Service Department	c9	49282
7166	214261	31-08-2014 01:36	Hired	Female	Service Department	c5	57742
7167	932441	31-08-2014 01:37	Hired	Male	Service Department	c5	69932
7168	39010	31-08-2014 01:38	Rejected	Male	Service Department	c5	14489
7169	686055	26-08-2014 12:14	Hired	Male	Operations Department	c5	54201
7170							=AVERAGE(G2:G7169)
7171							
7172							
7173							

Sheet1

Enter Accessibility: Good to go

Type here to search

Windows Start Taskbar: File Explorer, Edge, File, Home, Insert, Page Layout, Formulas, Data, Review, View, Help, Tell me what you want to do, Home tab ribbon, Font, Alignment, Number, Conditional Formatting, Format as Table, Cell Styles, Insert, Delete, Format, Cells, Editing, Sort & Find & Filter, Select, Share, Sheet1, +, 100%, 06:43 PM, 28-07-2022

Statistics - Excel

Aastha Kumar AK

Tell me what you want to do

File Home Insert Page Layout Formulas Data Review View Help

Font Alignment Number Conditional Formatting Format as Table Cell Styles Insert Delete Format Cells Editing

G7172 : fx

7152	641060	20-08-2014 17:49	Rejected	Male	Operations Department	i6	67779
7153	786222	20-08-2014 17:51	Hired	Male	Operations Department	i6	47286
7154	964206	25-08-2014 09:23	Hired	Male	Service Department	i5	7849
7155	586934	29-08-2014 19:00	Hired	Don't want to say	Service Department	c9	63159
7156	321440	17-08-2014 12:17	Hired	Female	Production Department	c9	89565
7157	975415	17-08-2014 07:48	Hired	Male	Service Department	c5	86162
7158	683982	17-08-2014 07:49	Hired	Don't want to say	Service Department	c5	58900
7159	665614	21-08-2014 07:30	Rejected	Male	Sales Department	c5	71449
7160	935497	27-08-2014 17:36	Rejected	Male	Service Department	c5	67196
7161	106032	22-08-2014 15:50	Hired	Male	Service Department	i7	16756
7162	166185	27-08-2014 05:01	Hired	Female	Service Department	i7	30952
7163	736189	28-08-2014 17:29	Hired	Male	Service Department	c9	64150
7164	614594	28-08-2014 17:30	Hired	Male	Service Department	c9	40152
7165	493131	28-08-2014 17:32	Hired	Male	Service Department	c9	49282
7166	214261	31-08-2014 01:36	Hired	Female	Service Department	c5	57742
7167	932441	31-08-2014 01:37	Hired	Male	Service Department	c5	69932
7168	39010	31-08-2014 01:38	Rejected	Male	Service Department	c5	14489
7169	686055	26-08-2014 12:14	Hired	Male	Operations Department	c5	54201
7170							49983.02902
7171							
7172							
7173							

Sheet1

Ready Accessibility: Good to go

Type here to search

Windows Start Taskbar: File Explorer, Edge, File, Home, Insert, Page Layout, Formulas, Data, Review, View, Help, Tell me what you want to do, Home tab ribbon, Font, Alignment, Number, Conditional Formatting, Format as Table, Cell Styles, Insert, Delete, Format, Cells, Editing, Sort & Find & Filter, Select, Share, Sheet1, +, 100%, 06:44 PM, 28-07-2022

Thus, the average salary is 49983.02902

C. Class Intervals: The class interval is the difference between the upper-class limit and the lower-class limit.

My task: Draw the class intervals for salary in the company?

D. Charts and Plots: This is one of the most important parts of analysis to visualize the data.

My task: Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working in different departments?

The screenshot shows a Microsoft Excel spreadsheet titled "Statistics - Excel". The "Insert" tab is active in the ribbon. A context menu is open over a selection of data in column D, showing options like "2-D Pie", "3-D Pie", "Doughnut", and "More Pie Charts...". The main data table has columns A through J, with data starting from row 7150. Column A contains Employee IDs, column B contains hire dates, column C contains departments, and column D contains genders. The "Service Department" is the largest category in the department column.

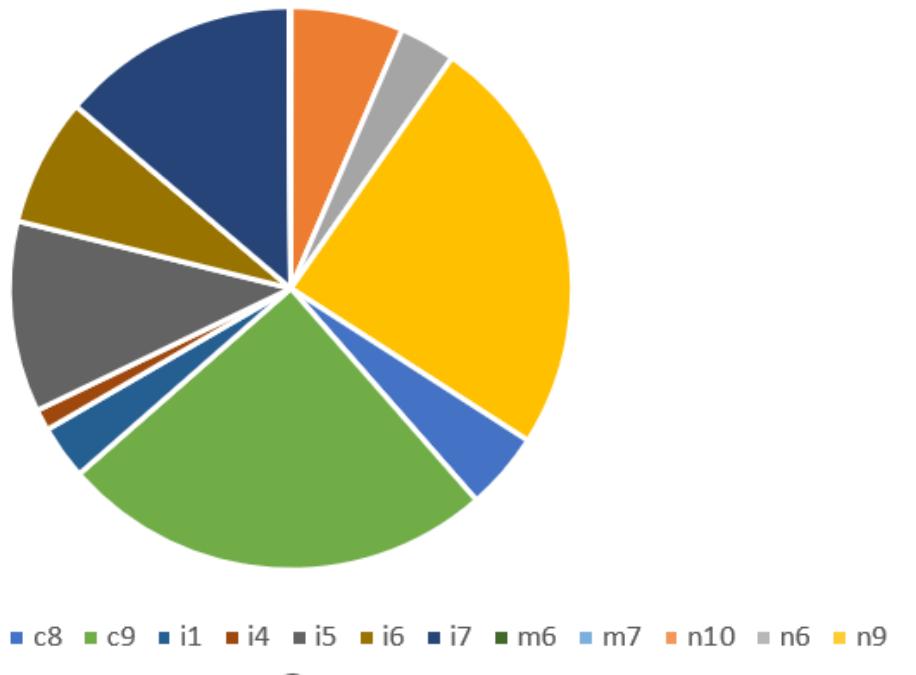


- | | |
|-----------------------------|-------------------------|
| ■ Finance Department | ■ General Department |
| ■ Human Resource Department | ■ Marketing Department |
| ■ Operations Department | ■ Production Department |
| ■ Purchase Department | ■ Sales Department |
| ■ Service Department | |

E. Charts: Use different charts and graphs to perform the task representing the data.

My task: Represent different post tiers using chart/graph?

The screenshot shows a Microsoft Excel spreadsheet titled "Statistics - Excel". The "Insert" tab is active in the ribbon. A dropdown menu for "Pie" charts is open, showing options for "2-D Pie", "3-D Pie", and "Doughnut". The main worksheet contains a table with columns: application_id, Interview Taken on, Status, event_name, Post Name, and Offered Salary. The "Offered Salary" column is sorted in descending order, with values ranging from 29462 to 56553. The "Post Name" column lists various department names. The status column shows entries like "Hired" and "Rejected". The "event_name" column includes entries such as "Male", "Female", and "-". The "Interview Taken on" column shows dates from May 2014. The "Status" column indicates the outcome of the interview.



Resources



XLS

[Statistics Dataset](#)

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

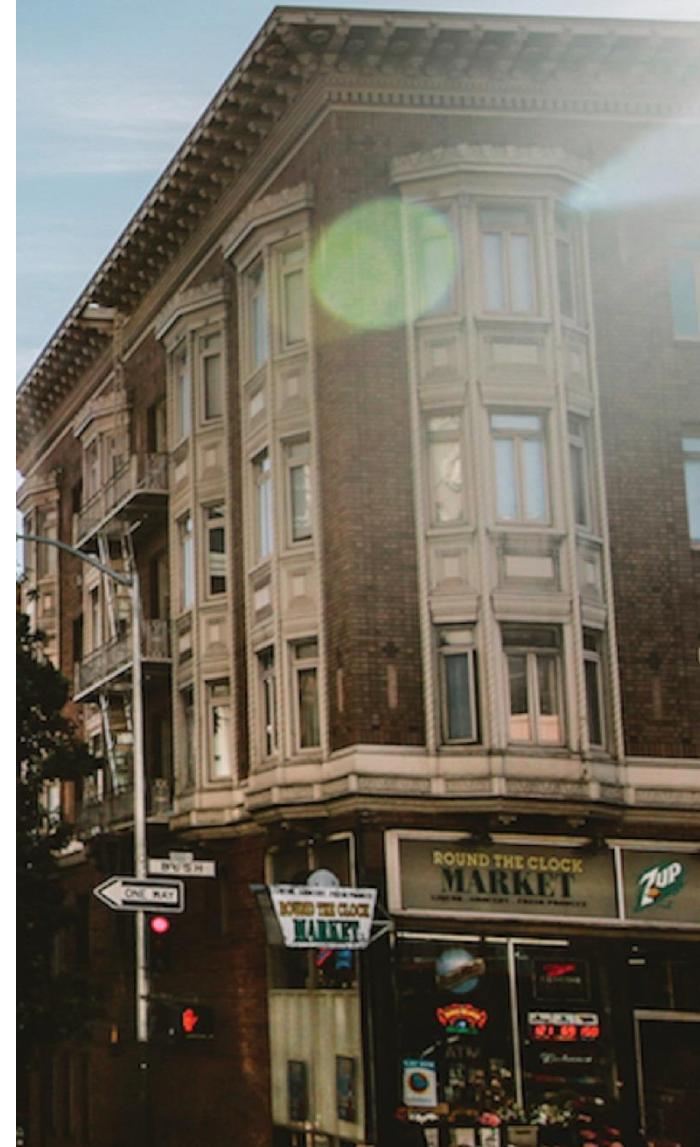
The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.

IMDB Movie Analysis

AUGUST 11

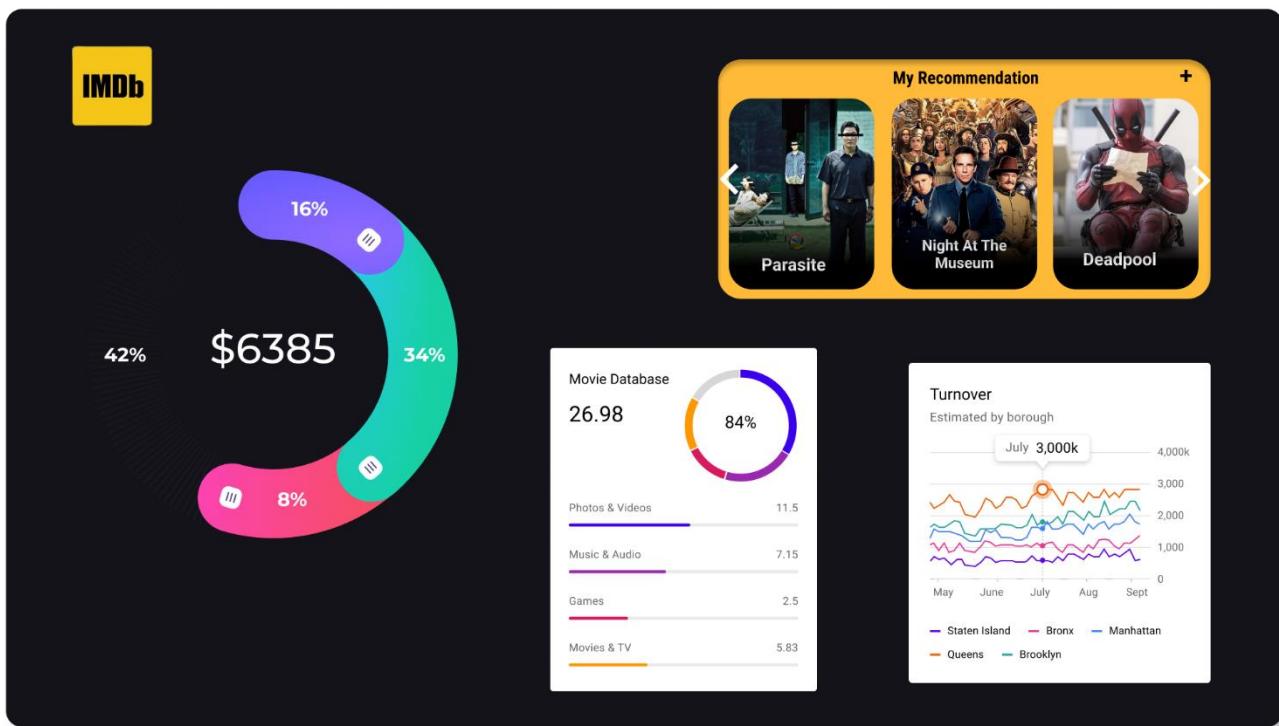
By Aastha Kumar



Description

I have been provided with a dataset having various columns of different IMDB Movies and am required to Frame the problem. For this task, I will need to define a problem I want to shed some light on.

I can do this by asking 'What?' This is where I frame the problem i.e. What is the problem? Once I have defined a problem I will use my Data Analysis skills to explore the data set and derive insights.



Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

Tech-Stack Used

- Google Sheet

Insights

A. Cleaning the data: This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

My Task: Clean the data

To delete all rows and columns of the spreadsheet, I used the keyboard shortcut **ctrl+A** to select all the values at once and then hit the delete button.

The screenshot shows a Microsoft Excel spreadsheet titled "IMDB_Movies - Excel". The spreadsheet contains a large dataset with approximately 22 columns and 30 rows of movie information. The columns include: color, director_n, num_critics, duration, director_f, factor_3, factor_2, n, actor_1, f, gross, genres, actor_1_n, movie_titl, num_vote, cast_total, actor_3_n, face, num_plot, keyw, movie_imdb, num_user, lar, and others. The data includes various movie titles like "Avatar", "The Dark Knight", "Star Wars", and "The Avengers", along with their respective actors, directors, and box office numbers. The Excel interface shows the Home tab selected, with various tools and ribbon tabs visible at the top. The status bar at the bottom indicates the average value is 4871785.233, there are 138535 counts, and the sum is 3.83064E+11. The date and time shown are 03-08-2022 at 12:46 PM.

color	director_n	num_critics	duration	director_f	factor_3	factor_2	n	actor_1	f	gross	genres	actor_1_n	movie_titl	num_vote	cast_total	actor_3_n	face	num_plot	keyw	movie_imdb	num_user	lar
Color	James Cam	723	178	0	855	Joel David	1000	7.61E+08	Action Ad	CCH Pouni Avatar	886204	4834	Wes Studi	0	avatar fut	http://ww	3054	En				
Color	Gore Verb	302	169	563	1000	Orlando B	40000	3.09E+08	Action Ad	Johnny De Pirates	471220	48350	Jack Davel	0	goddes r	http://ww	1238	En				
Color	Sam Menc	602	148	0	161	Rory Kinne	11000	2E+08	Action Ad	Christoph Spectre	275868	11700	Stephanie	1	bomb esp	http://ww	994	En				
Color	Christoph	813	164	22000	23000	Christian E	27000	4.48E+08	Action Th	Tom Hard The Dark I	1144337	106759	Joseph Go	0	deception h	http://ww	2701	En				
Doug Walker				131		Rob Walk	131		Document Dou	Wall Star Wars:	8	143		0		http://ww						
Color	Andrew St	462	132	475	530	Samantha	640	73058679	Action Ad	Daryl Sabz John Carte	212204	1873	Polly Walk	1	alien jame	http://ww	738	En				
Color	Sam Raim	392	156	0	4000	James Fra	24000	3.37E+08	Action Ad	J.K. Simmc Spider-Ma	383056	46055	Kirsten Du	0	sandman r	http://ww	1902	En				
Color	Nathan Gr	324	100	15	284	Donna Mt	799	2.01E+08	Action Ad	Bra Brad Garr Tangled	294810	2036	M.C. Gain	1	17th cent t	http://ww	387	En				
Color	Joss Whec	635	141	0	19000	Robert Do	26000	4.59E+08	Action Ad	Chris Hem Avengers:	462669	92000	Scarlett Jo	4	artificial i h	http://ww	1117	En				
Color	David Yate	375	153	282	10000	Daniel Rac	25000	3.02E+08	Action Ad	Alan Rickn Harry Pott	321795	58753	Rupert Gri	3	blood bo	http://ww	973	En				
Color	Zack Snyd	673	183	0	2000	Lauren Co	15000	3.3E+08	Action Ad	Henry Cav Batman v	371639	24450	Alan D. Pu	0	based on r	http://ww	3018	En				
Color	Bryan Sing	434	169	0	903	Marlon Br	18000	2E+08	Action Ad	Kevin Spa Superman	240396	29991	Frank Lanq	0	crystal ep	http://ww	2367	En				
Color	Marc Fors	403	106	395	393	Mathieu A	451	1.68E+08	Action Ad	Giancarlo Quantum	330784	2023	Rory Kinne	1	action her h	http://ww	1243	En				
Color	Gore Verb	313	151	563	1000	Orlando B	40000	4.23E+08	Action Ad	Johnny De Pirates of	522040	48486	Jack Davel	2	box office h	http://ww	1832	En				
Color	Gore Verb	450	150	563	1000	Ruth Wilsc	40000	89289910	Action Ad	Johnny De The Lone I	181792	45757	Tom Wilki	1	horse out h	http://ww	711	En				
Color	Zack Snyd	733	143	0	748	Christoph	15000	2.91E+08	Action Ad	Henry Cav Man of St	548573	20495	Harry Leni	0	based on r	http://ww	2536	En				
Color	Andrew Ai	258	150	80	201	Pierfrance	22000	1.42E+08	Action Ad	Peter Dink The Chron	149922	22697	Damiān J	4	brother br h	http://ww	438	En				
Color	Joss Whec	703	173	0	19000	Robert Do	26000	6.23E+08	Action Ad	Chris Hem The Aveng	995415	87697	Scarlett Jo	3	alien inva h	http://ww	1722	En				
Color	Rob Marsl	448	136	252	1000	Sam Claflii	40000	2.41E+08	Action Ad	Johnny De Pirates of	370704	54083	Stephen G	4	blackbear r	http://ww	484	En				
Color	Barry Soni	451	106	188	718	Michael St	10000	1.79E+08	Action Ad	Will Smith Men in Bl	268154	12572	Nicole Sch	1	alien crim h	http://ww	341	En				
Color	Barry Soni	427	164	0	773	Adam Rec	5000	2.55E+08	Action Ad	Aidan Tora Tho Nohki	254270	9153	Liam Ne	0	removal f	http://ww	901	En				

IMDB_Movies - Excel

Aastha Kumar AK Share

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Font Alignment Number Styles Cells

Clipboard Paste Merge & Center Conditional Formatting Table Styles Insert Delete Format Sort & Find & Filter Select

A1 A B C D E F G H I J K L M N O P Q R S

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

IMDB_Movies Ready Accessibility: Unavailable

Type here to search

Windows Start button Taskbar icons Network connection battery level Date and time

B. Movies with highest profit: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.

My Task: Find the movies with the highest profit.

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	actor_1_n	movie_titl	num_vote	cast_total	actor_3_n	facenumb	plot_keyw	movie_im	num_user	language	country	content_r	budget	title_year	actor_2_f	imdb_sco	aspect_ra	movie_fac	profit
2	CCH Poun	Avatar	886204	4834	Wes Studi	0	avatar fut	http://ww	3054	English	USA	PG-13	2.37E+08	2009	936	7.9	1.78	33000	=I2-W2
3	Ishaaan De	Director of	471220	48250	Lark Davis	0	anddacelebr	http://ww	1228	English	INDIA	PG-12	2.2E+08	2007	500	7.1	2.35	33000	

IMDB Movies - Excel

Aastha Kumar AK

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

AC1 profit

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	num_vote	cast_total	actor_3_n	facenumb	plot_keyw	movie_im	num_user	language	country	content_r	budget	title_year	actor_2_f	imdb_sco	aspect_ra	movie_fac	profit		
2	886204	4834	Wes Studi	0	avatar fut	http://ww	3054	English	USA	PG-13	2.37E+08	2009	936	7.9	1.78	33000	523505847		
3	471220	48350	Jack Dauer	0	goddess r	http://ww	1238	English	USA	PG-13	3E+08	2007	5000	7.1	2.35	0	9404152		
4	275868	11700	Stephanie	1	bomb esp	http://ww	994	English	UK	PG-13	2.45E+08	2015	393	6.8	2.35	85000	-44925825		
5	1144337	106759	Joseph Go	0	deception	http://ww	2701	English	USA	PG-13	2.5E+08	2012	23000	8.5	2.35	164000	198130642		
6	8	143	0	http://ww									12	7.1	0	0	0	0	
7	212204	1873	Polly Walk	1	alien ame	http://ww	738	English	USA	PG-13	2.64E+08	2012	632	6.6	2.35	24000	-1.9E+08		
8	383056	46055	Kirsten Du	0	sandman	http://ww	1902	English	USA	PG-13	2.58E+08	2007	11000	6.2	2.35	0	78530303		
9	294810	2036	M.C. Gain	1	17th centu	http://ww	387	English	USA	PG	2.6E+08	2010	553	7.8	1.85	29000	-59192738		
10	462669	92000	Scarlett Jo	4	artificial i	http://ww	1117	English	USA	PG-13	2.5E+08	2015	21000	7.5	2.35	118000	208991599		
11	321795	58753	Rupert Gri	3	blood bo	http://ww	973	English	UK	PG	2.5E+08	2009	11000	7.5	2.35	10000	51956980		
12	371639	24450	Alan D. Pu	0	based on	http://ww	3018	English	USA	PG-13	2.5E+08	2016	4000	6.9	2.35	197000	80249062		
13	240396	29991	Frank Lang	0	crystal ep	http://ww	2367	English	USA	PG-13	2.09E+08	2006	10000	6.1	2.35	0	-8930592		
14	330784	2023	Rory Kinne	1	action her	http://ww	1243	English	UK	PG-13	2E+08	2008	412	6.7	2.35	0	-31631573		
15	522040	48486	Jack Dauer	2	box office	http://ww	1832	English	USA	PG-13	2.25E+08	2006	5000	7.3	2.35	5000	198032628		
16	181792	45757	Tom Wilki	1	horse out	http://ww	711	English	USA	PG-13	2.15E+08	2013	2000	6.5	2.35	48000	-1.26E+08		
17	548573	20495	Harry Leni	0	based on	http://ww	2536	English	USA	PG-13	2.25E+08	2013	3000	7.2	2.35	118000	66021565		
18	149922	22697	Damiān A	4	brother br	http://ww	438	English	USA	PG	2.25E+08	2008	216	6.6	2.35	0	-83385977		
19	995415	87697	Scarlett Jo	3	alien inva	http://ww	1722	English	USA	PG-13	2.2E+08	2012	21000	8.1	1.85	123000	403279547		
20	370704	54083	Stephen G	4	blackbear	http://ww	484	English	USA	PG-13	2.5E+08	2011	11000	6.7	2.35	58000	-8936125		
21	268154	12572	Nicole Sch	1	alien crim	http://ww	341	English	USA	PG-13	2.25E+08	2012	816	6.8	1.85	40000	-45979146		
22	254739	0153	Immae Nor	0	remulaf	http://ww	902	English	New Zealand	PG-12	7.5E+08	2014	672	7.5	2.35	65000	5109270		

IMDB_Movies

Ready Accessibility:Unavailable Average: 4097943.942 Count: 5044 Sum: 20665931298

Type here to search

Windows 10 01:03 PM 03-08-2022

C. Top 250: Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than

25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.

My Task: Find IMDB Top 250

Select the entire spreadsheet by hitting **ctrl+A**

The screenshot shows an Excel spreadsheet titled "IMDB_Movies". The table has approximately 250 rows and 25 columns. The columns include: movie_id, title, num_votes, cast_total, actor_3_name, facenumber, plot_keywords, movie_imdb_link, num_user_reviews, language, country, content_rating, budget, title_year, actor_2_name, imbd_score, aspect_ratio, movie_faculty, profit, and IMDb_Top. The "IMDb_Top" column contains numerical values ranging from 0 to 250. The "IMDb_Top" column is located at the end of each row, indicating the rank of the movie in the top 250. The data is presented in a tabular format with various movie details such as plot, year, and financial information.

Under Sort and Filter menu, click the custom sort option

The screenshot shows the Excel ribbon with the "Sort & Filter" icon selected. A "Sort" dialog box is open, showing the "Custom Sort..." option selected. The "Sort by" dropdown is set to "imdb_score", "Sort On" is "Cell Values", and "Order" is "Largest to Smallest". The "My data has headers" checkbox is checked. The main spreadsheet area shows a table of movie data with columns for title, plot, year, and various metrics like budget and rating. The "IMDb_Top" column is present at the end of each row, indicating the rank of the movie in the top 250. The data is presented in a tabular format with various movie details such as plot, year, and financial information.

Click OK. All the contents of the spreadsheet will be sort according to the highest to lowest imdb_score. Select the top 250 movies corresponding to the highest imdb_scores and paste the values in the new column created (IMDb_Top_250)

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	plot_keywords	num_userReviews	language	country	content_rating	budget	title_year	actor_2_imdb_score	aspect_ratio	movie_faculty	profit	IMDb_Top_250							
2	http://www.imdb.co	English	Canada					179	9.5	1.33	0	0	Towering Inferno						
3	escape frch	http://ww	4144	English	USA	R	25000000	1994	745	9.3	1.85	108000	3341469	The Shawshank Redemption					
4	crime fam	http://ww	2238	English	USA	R	6000000	1972	10000	9.2	1.85	43000	1.29E+08	The Godfather					
5	meaning ch	http://ww	37	Polish	Poland	TV-MA			3	9.1	1.33	0	447093	Dekalog					
6	meaning cht	http://ww	37	Polish	Poland	TV-MA			3	9.1	1.33	0	447093	Dekalog					
7	http://ww	1	USA		17000000	2016	454	9.1		0	-1.7E+07	Kickboxer: Vengeance							
8	based on	http://ww	4667	English	USA	PG-13	1.85E+08	2008	13000	9	2.35	37000	3.48E+08	The Dark Knight					
9	1950s cor	http://ww	650	English	USA	R	13000000	1974	14000	9	1.85	14000	44300000	The Godfather: Part II					
10	anthology	http://ww	173	English	USA	TV-MA			1000	9	1.78	61000	0	Fargo					
11	battle lepi	http://ww	3189	English	USA	PG-13	94000000	2003	857	8.9	2.35	16000	2.83E+08	The Lord of the Rings: The Return of the King					
12	german jg	http://ww	1273	English	USA	R	22000000	1993	795	8.9	1.85	41000	74067179	Schindler's List					
13	black com	http://ww	2195	English	USA	R	8000000	1994	902	8.9	2.35	45000	99930000	Pulp Fiction					
14	civil war h	http://ww	780	Italian	Italy	Approved	1200000	1966	34	8.9	2.35	20000	4900000	The Good, the Bad and the Ugly					
15	courtroon	http://ww	888	English	USA	Not Rated	350000	1957	259	8.9	1.66	40000	-350000	12 Angry Men					
16	ambiguou	http://ww	2803	English	USA	PG-13	1.6E+08	2010	27000	8.8	2.35	175000	1.33E+08	Inception					
17	elf hobbit	http://ww	5060	English	New Zealand	PG-13	93000000	2001	5000	8.8	2.35	21000	2.21E+08	The Lord of the Rings: The Fellowship of the Ring					
18	corruptionr	http://ww	394	English	USA	TV-MA			4	8.8	16	55000	0	Daredevil					
19	anti estab	http://ww	2968	English	USA	R	63000000	1999	783	8.8	2.35	48000	-2.6E+07	Fight Club					
20	amputeej	http://ww	1398	English	USA	PG-13	55000000	1994	294	8.8	2.35	59000	2.75E+08	Forrest Gump					
21	alcoholib	http://ww	151	English	USA	TV-MA			547	8.8	1.33	0	0	It's Always	Philadelphia				
22	dwellfrom	http://ww	900	English	USA	PG	18000000	1980	504	9	2.25	17000	2.72E+08	Star Wars	(Ctrl)	The Empire Strikes Back			

D. Best Directors: TGroup the column using the director_name column.

Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

My Task: Find the best directors

E. Popular Genres: Perform this step using the knowledge gained while performing previous steps.

My Task: Find popular genres

F. Charts: Create three new columns namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined. Group the combined column using the actor_1_name column. Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean. Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on

the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df_by_decade.

My Task: Find the critic-favorite and audience-favorite actors

Resources



[IMDB Movies](#)

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

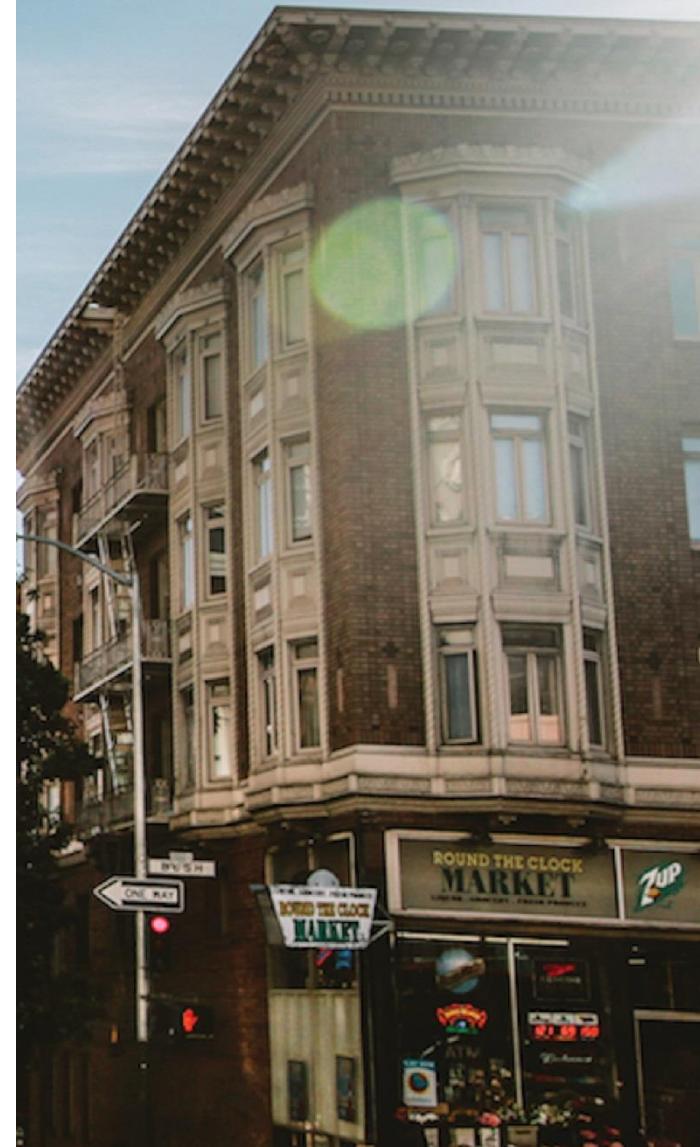
The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.

Bank Loan Case Study

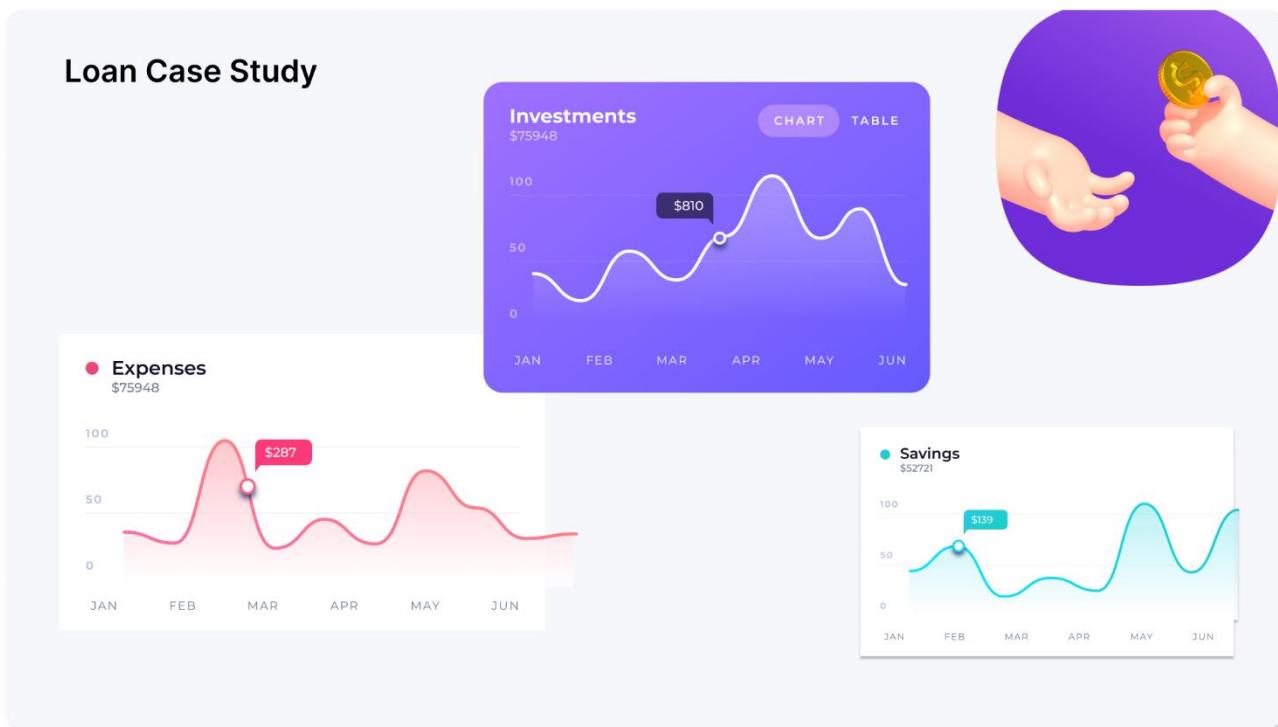
AUGUST 11

By Aastha Kumar



Description

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.



Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

Tech-Stack Used

- Google Sheet

Insights

- A. Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)
- B. Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier. Again, remember that for this exercise, it is not necessary to remove any data points.
- C. Identify if there is data imbalance in the data. Find the ratio of data imbalance.

Hint: Since there are a lot of columns, you can run your analysis in loops for the appropriate columns and find the insights.

- D. Explain the results of univariate, segmented univariate, bivariate analysis, etc. in business terms.
- E. Find the top 10 correlation for the Client with payment difficulties and all other cases (Target variable). Note that you have to find the top correlation by segmenting the data frame w.r.t to the target variable and then find the top correlation for each of the segmented data and find if any insight is there. Say, there are 5+1(target) variables in a dataset: Var1, Var2, Var3, Var4, Var5, Target. And if you have to find top 3 correlation, it can be: Var1 & Var2, Var2 & Var3, Var1 & Var3. Target variable will not feature in this correlation as it is a categorical variable and not a continuous variable which is increasing or decreasing.
- F. Include visualizations and summarize the most important results in the presentation. You are free to choose the graphs which explain the numerical/categorical variables. Insights should explain why the variable is

important for differentiating the clients with payment difficulties with all other cases.

Resources



[Loan Case Study](#)

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.

ABC CALL VOLUME TREND ANALYSIS

SEPTEMBER 26
By Aastha Kumar



Description

For your final project we are providing you with a dataset of a Customer Experience (CX) Inbound calling team for 23 days. Data includes Agent_Name, Agent_ID, Queue_Time [duration for which customer have to wait before they get connected to an agent], Time [time at which call was made by customer in a day], Time_Bucket [for easiness we have also provided you with the time bucket], Duration [duration for which a customer and executives are on call, Call_Seconds [for simplicity we have also converted those time into seconds], call status (Abandon, answered, transferred).

A customer experience (CX) team consists of professionals who analyze customer feedback and data, and share insights with the rest of the organization. Typically, these teams fulfil various roles and responsibilities such as: Customer experience programs (CX programs), Digital customer experience, Design and processes, Internal communications, Voice of the customer (VoC), User experiences, Customer experience management, Journey mapping, Nurturing customer interactions, Customer success, Customer support, Handling customer data, Learning about the customer journey.

Let's look at some of the most impactful AI-empowered customer experience tools you can use today:

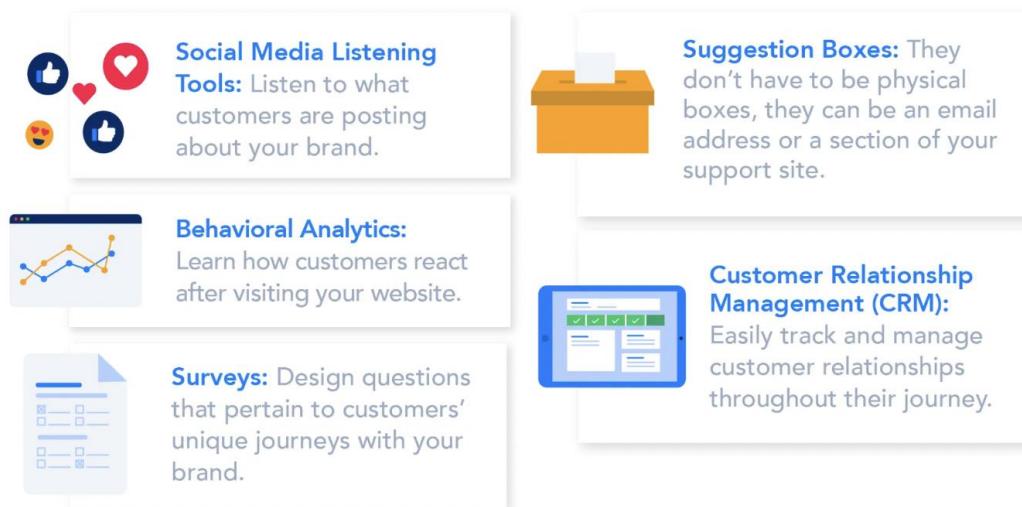
Interactive Voice Response (IVR), Robotic Process Automation (RPA), Predictive Analytics, Intelligent Routing

In a Customer Experience team there is a huge employment opportunities for Customer service representatives A.k.a. call centre agents, customer service agents. Some of the roles for them include: Email support, Inbound support, Outbound support, social media support.

Inbound customer support is defined as the call centre which is responsible for handling inbound calls of customers. Inbound calls are the incoming voice calls of the existing customers or prospective customers for your business which are attended by customer care representatives. Inbound customer service is the methodology of attracting, engaging, and delighting your customers to turn them into your business' loyal advocates. By solving your customers' problems and helping them achieve

success using your product or service, you can delight your customers and turn them into a growth engine for your business.

Tools to Optimize Your Customer Experience



Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

Tech-Stack Used

- Google Sheet

Insights

- a) Calculate the average call time duration for all incoming calls received by agents (in each Time_Bucket).

Use the Filter option to filter out only one slot (say 9_10) from the Time_Bucket column. Now all other fields except 9_10 is hidden.

The screenshot shows a Microsoft Excel spreadsheet titled "Call_Volume_Trend_Analysis_Project_9 - Excel". The data is organized into columns: Agent_Na, Agent_ID, Customer, Queue_Ti, Date_&_T, Time, Time_Buc, Duration, Call_Sec, Call_Status, Wrapped, Ringing, and IVR_Dura. A context menu is open over the "Time_Buc" column header, specifically over the cell containing "9_10". The menu options include "Sort A to Z", "Sort Z to A", "Sort by Color", "Sheet View", "Clear Filter From 'Time_Bucket'", "Filter by Color", and "Text Filters". A sub-menu under "Text Filters" is open, showing a search field and a list of numerical ranges: 14_15, 15_16, 16_17, 17_18, 18_19, 19_20, and 20_21. The cell "9_10" is checked. At the bottom of the filter menu, there are "OK" and "Cancel" buttons. The status bar at the bottom right shows the time as 11:38 AM and the date as 26-09-2022.

Calculate the average call time duration by using the simple formula:
Average = sum of all observations / number of observations

Call_Volume_Trend_Analysis_Project_9 - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Agent_Na	Agent_ID	Customer_	Queue_Ti	Date_&_Ti	Time	Time_Buc	Duration(I	Call_Secon	Call_Status	Wrapped	Ringing	IVR_Dura	n
2	Executives 4:	1000042	98502XXXXX	2	01-01-2022	9.00 9_10		00:01:36	96.00	answered	Agent	YES	00:00:16	
3	Executives 4	1000004	80595XXXXX	0	01-01-2022	9.00 9_10		00:02:20	140.00	answered	Agent	YES	00:00:26	
4	Executives 6:	1000065	70202XXXXX	0	01-01-2022	9.00 9_10		00:01:25	85.00	answered	AutoWrappe	YES	00:00:16	
5	Executives 5:	1000055	96104XXXXX	1	01-01-2022	9.00 9_10		00:01:31	91.00	answered	Agent	YES	00:00:25	
6	Executives 2:	1000021	82001XXXXX	0	01-01-2022	9.00 9_10		00:02:45	165.00	answered	Agent	YES	00:00:23	
7	#N/A	#N/A	96424XXXXX	13	01-01-2022	9.00 9_10		00:00:00	0.00	abandon		YES	00:00:16	
8	Executives 5!	1000055	96737XXXXX	79	01-01-2022	9.00 9_10		00:01:25	85.00	answered	AutoWrappe	YES	00:00:13	
9	#N/A	#N/A	96392XXXXX	60	01-01-2022	9.00 9_10		00:00:00	0.00	abandon		YES	00:00:17	
10	Executives 4:	1000042	90820XXXXX	52	01-01-2022	9.00 9_10		00:01:05	65.00	answered	Agent	YES	00:00:20	
11	Executives 6:	1000065	97410XXXXX	62	01-01-2022	9.00 9_10		00:03:00	180.00	answered	AutoWrappe	YES	00:00:44	
12	Executives 4	1000040	70076XXXXX	52	01-01-2022	9.00 9_10		00:01:48	108.00	answered	Agent	YES	00:00:15	
13	Executives 2:	1000021	82505XXXXX	89	01-01-2022	9.00 9_10		00:03:06	186.00	answered	Agent	YES	00:00:16	
14	#N/A	#N/A	97232XXXXX	120	01-01-2022	9.00 9_10		00:00:00	0.00	abandon		YES	00:00:40	
15	Executives 5!	1000055	96392XXXXX	45	01-01-2022	9.00 9_10		00:01:40	100.00	answered	AutoWrappe	YES	00:00:42	
16	Executives 4:	1000042	97471XXXXX	55	01-01-2022	9.00 9_10		00:01:15	75.00	answered	AutoWrappe	YES	00:00:19	
17	#N/A	#N/A	77082XXXXX	16	01-01-2022	9.00 9_10		00:00:00	0.00	abandon		YES	00:00:18	
18	#N/A	#N/A	95255XXXXX	44	01-01-2022	9.00 9_10		00:00:00	0.00	abandon		YES	00:00:17	
19	Executives 4	1000004	79725XXXXX	88	01-01-2022	9.00 9_10		00:04:03	243.00	answered	AutoWrappe	YES	00:00:15	
20	Executives 4!	1000049	98344XXXXX	46	01-01-2022	9.00 9_10		00:04:10	250.00	answered	Agent	YES	00:00:19	

115048	Executives 6!	1000065	79813XXXXX	7	23-01-2022	9.00 9_10		00:05:58	358.00	answered	Agent	YES	00:00:17
117990								00:02:19					
117991													

Repeat the same with all other baskets.

- b) Show the total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time]. You can select time in a bucket form (i.e. 1-2, 2-3,)
- c) As you can see current abandon rate is approximately 30%. Propose a manpower plan required during each time bucket [between 9am to 9pm] to reduce the abandon rate to 10%. (i.e. You have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.)
- d) Let's say customers also call this ABC insurance company in night but didn't get answer as there are no agents to answer, this creates a bad customer experience for this Insurance company. Suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] and distribution of those 30 calls are as follows:

Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)													
9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am		
3	3	2	2	1	1	1	1	3	4	4	5		

Now propose a manpower plan required during each time bucket in a day.
Maximum Abandon rate assumption would be same 10%.

Resources



Call Volume Trend Analysis

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.