

# CS520 - Data Integration, Warehousing and Provenance

## Data Curation Project

### **Guided By -**

Dr. Boris Glavic  
Pengyuan Li (TA)

### **Presented By -**

Purva Lila (A20491945)  
Shubham Singh (A20492876)  
Suraj Nammi (A20501709)

# Dataset

FIFA DATASET (SPORTS DOMAIN) - 18980 records

<https://www.kaggle.com/yagunnersya/fifa-21-messy-raw-dataset-for-cleaning-exploring>

	Original DataSet	Modified DataSet
Number of Attributes	77	24
Size	16.3 MB	4.82 MB
Data Format	CSV	CSV

Column reduction due to:

- Increased Data Size
- No curation tasks in the removed columns
- Only numeric values and all clean values

# Sample DataSet

ID	Name	LongName	photoU	playerU	Nation	Age	â†“O	P	Cl	Contract	Position	Height	Weight	Preferre	BC	Best P	Joined	Loan De	Value	Wage	Release	A/W	D/W
194222	Laguardia	VÅ-ctor Laguardia	https://cdi	http://sofi	Spain	30	78	78		2014 ~ 2022	CB	6'1"	176lbs	Right	78	CB	11-Jul-14		â~8.5M	â~25K	â~17.9M	Medium	High
192546	E. Stevens	Enda Stevens	https://cdi	http://sofi	Republic o	29	78	78		2017 ~ 2023	LWB, LB	6'0"	172lbs	Left	78	LWB	1-Jul-17		â~8.5M	â~37K	â~16.2M	Medium	Medium
188427	S. LuliÄ†	Senad LuliÄ†	https://cdi	http://sofi	Bosnia Her	34	78	78		2011 ~ 2021	LM	6'0"	172lbs	Right	78	LM	1-Jul-11		â~5M	â~56K	â~8.5M	High	High
183353	VÅ-ctor DÅ-az	VÅ-ctor David DÅ	https://cdi	http://sofi	Spain	32	78	78		2017 ~ 2022	RB	6'0"	159lbs	Right	78	RB	4-Jul-17		â~6.5M	â~26K	â~13.7M	Medium	High
154472	Jorge Molina	Jorge Molina Vida	https://cdi	http://sofi	Spain	38	78	78		2020 ~ 2022	ST	6'2"	159lbs	Right	78	ST	#####		â~3.7M	â~24K	â~7.8M	Medium	Medium
207494	J. Lingard	Jesse Lingard	https://cdi	http://sofi	England	27	77	77		2011 ~ 2021	CAM, RM	175cm	62kg	Right	77	CAM	1-Jul-11		â~11M	â~95K	â~18.1M	High	High
158963	L. Biglia	Lucas Biglia	https://cdi	http://sofi	Argentina	34	77	77		2020 ~ 2021	CDM, CM	178cm	63kg	Right	77	CDM	#####		â~3.6M	â~15K	â~6.8M	Medium	Medium
226370	AndrÅ© Horta	AndrÅ© Filipe Luz	https://cdi	http://sofi	Portugal	23	75	82		2019 ~ 2024	CM	175cm	62kg	Right	77	CAM	1-Jul-19		â~11M	â~12K	â~19.8M	High	Medium
251804	S. Dest	SergiÅ±o Dest	https://cdi	http://sofi	United Sta	19	75	86		2020 ~ 2025	RB	175cm	62kg	Right	77	RM	1-Oct-20		â~11.5M	â~50K	â~22.5M	High	Low
225706	M. Trauco	Miguel Trauco	https://cdi	http://sofi	Peru	27	74	74		2019 ~ 2022	LB, LWB, L	175cm	62kg	Left	74	LB	6-Aug-19		â~4.1M	â~24K	â~10M	Medium	Medium
221282	J. Stacey	Jack Stacey	https://cdi	http://sofi	England	24	73	78		2019 ~ 2023	RWB, RB	178cm	62kg	Right	73	RWB	8-Jul-19		â~4.3M	â~10K	â~9.3M	High	Medium
210831	B. Verstraete	Birger Verstraete	https://cdi	http://sofi	Belgium	26	72	76		Jun 30, 2021	CDM, CM	175cm	61kg	Right	73	CDM	1-Jul-19	30-Jun-21	â~2.9M	â~18K	â~0	Medium	Medium
253157	H. Boudaoui	Hichem Boudaoui	https://cdi	http://sofi	Algeria	20	72	82		2019 ~ 2023	CM, RM	175cm	61kg	Right	74	CAM	2-Sep-19		â~5M	â~17K	â~10.5M	High	High
239475	J. GonzÅ±lez	Jonathan GonzÅ±l	https://cdi	http://sofi	Mexico	21	70	80		2017 ~ 2022	CM, CDM	175cm	62kg	Right	73	CDM	1-Jul-17		â~3.6M	â~16K	â~5.3M	Medium	High
233409	R. Gumny	Robert Gumny	https://cdi	http://sofi	Poland	22	70	81		2020 ~ 2025	RB	178cm	61kg	Right	71	RWB	2-Sep-20		â~3.4M	â~12K	â~5.6M	High	High
221875	Son Jun Ho	Jun Ho Son	https://cdi	http://sofi	Korea Rep	28	70	70		2018 ~ 2025	CM, CDM	178cm	62kg	Right	70	CM	7-Jan-18		â~1.6M	â~7K	â~2.3M	Medium	Medium
227900	N. Tripichio	NicolÅ±s Tripichio	https://cdi	http://sofi	Argentina	24	69	72		2018 ~ 2021	RB, RM	175cm	61kg	Right	70	RWB	23-Jul-18		â~1.8M	â~6K	â~1.9M	Medium	Medium
183593	J. John	Joshua John	https://cdi	http://sofi	Aruba	31	68	68		2020 ~ 2021	LW	175cm	63kg	Right	68	LW	4-Sep-20		â~1M	â~4K	â~1.1M	Medium	Low
258922	Marvin	Marvin Olawale A	https://cdi	http://sofi	Spain	20	68	80		2019 ~ 2021	RW	178cm	63kg	Right	68	RW	1-Jul-19		â~2.9M	â~36K	â~3.6M	High	Low
230357	Kaio Aldair	Kaio Jesus Aldair F	https://cdi	http://sofi	Brazil	28	68	68		2019 ~ 2023	RB, RM	175cm	63kg	Right	68	RB	1-Jan-19		â~1.1M	â~6K	â~1.5M	High	Medium
224798	R. Gall	Romain Gall	https://cdi	http://sofi	United Sta	25	66	69		Dec 31, 2020	LM	175cm	63kg	Right	67	CAM	18-Jul-18	#####	â~1.1M	â~3K	â~0	High	Low
253008	K. Å±lvarez	Kevin Å±lvarez	https://cdi	http://sofi	Mexico	21	65	74		2018 ~ 2022	RB, RM	175cm	63kg	Right	65	RB	1-Jan-18		â~1.5M	â~3K	â~1.4M	High	Medium
220544	J. Arrascaita	Jaime Arrascaita	https://cdi	http://sofi	Bolivia	26	65	65		2020 ~ 2021	LM	175cm	62kg	Right	65	LM	1-Jan-20		â~775K	â~500	â~1.2M	Medium	Medium
156321	A. Akinfenwa	Adebayo Akinfenwa	https://cdi	http://sofi	England	38	65	65		2016 ~ 2021	ST	178cm	110kg	Right	65	ST	10-Jul-16		â~240K	â~3K	â~352K	Low	Low
219948	J. Cummings	Jason Cummings	https://cdi	http://sofi	Scotland	24	65	69		2019 ~ 2021	ST, RW	178cm	63kg	Left	67	ST	2-Sep-19		â~1M	â~3K	â~1.3M	Medium	Medium
209526	S. Kavanagh	Sean Kavanagh	https://cdi	http://sofi	Republic o	26	64	67		2018 ~ 2020	LB, LM	175cm	62kg	Left	65	CAM	#####		â~700K	â~900	â~813K	High	Medium
242841	A. Appiah	Arvin Appiah	https://cdi	http://sofi	England	19	64	84		2019 ~ 2024	LM, RM	175cm	63kg	Left	65	LW	2-Sep-19		â~1.6M	â~1K	â~2.5M	High	High

## **Overall Curation Tasks in the Dataset**

- Empty Values (Missing Values)
- Wrong Values (Special characters, Non Ascii characters , Copyright\_symbol , Trademark Symbol)
- Irrelevant column header
- Standardize K and M
- Standardize feet and cms
- Standardize KG and LBS
- Standardize Date formats
- Inter-column relationship - Contract and Loan

## Issues in the Dataset

1. **Name** contains special characters (Non Ascii characters , Copyright\_symbol , Trademark Symbol )
2. **Long name** contains Double quotes, non ascii characters, symbols (%), Copyright symbol
3. **Nationality** contains Copyright symbol, non Ascii characters.

ID	Name	LongName
158023	L. Messi	Lionel Messi
231747	K. Mbapp��	Kylian Mbapp��
192448	M. ter Stegen	Marc-Andr�� ter Stegen
203376	V. van Dijk	Virgil van Dijk
208722	S. Man��	Sadio Man��
200145	Casemiro	Carlos Henrique Venancio Casimiro
165153	K. Benzema	Karim Benzema
155862	Sergio Ramos	Sergio Ramos Garc��a
153079	S. Ag��ero	Sergio Ag��ero
202652	R. Sterling	Raheem Sterling
215914	N. Kant��	N'Golo Kant��
212622	J. Kimmich	Joshua Kimmich

Nationality
Albania
Albania
S��o Tom�� & Pr��ncipe
Albania
Albania
Afghanistan
Albania
Albania

## Issues in the Dataset continued...

4. **Value** contains non Ascii character , standardize K and M
5. **Wage** contains non Ascii character , standardize K and M
6. **Release Clause** contains non Ascii character , standardize K and M

Value	Wage	Release Clause
â,-0	â,-0	â,-0
â,-0	â,-0	â,-0
â,-110K	â,-1K	â,-75K
â,-110K	â,-1K	â,-75K
â,-110K	â,-2K	â,-109K
â,-110K	â,-2K	â,-109K
â,-1.1M	â,-13K	â,-2.2M
â,-1.1M	â,-950	â,-3.5M
â,-1.1M	â,-850	â,-3.3M
â,-0	â,-0	â,-0

## Issues in the Dataset continued...

7. **Contract** contains Free. Loan, time-date different representation format
8. **Loan** contains null fields (**CONTRACT (not under loan) AND LOAN (loan = null) RELATION**)
9. Standardize Date Formats

Contract	Joined	Loan Date End
Free	10-Aug-18	
Free	10-Aug-18	
2020 ~ 2027	16-Sep-20	
Dec 31, 2021 On Loan	9-Jul-16	31-Dec-21
Free	1-Jul-15	
Dec 31, 2021 On Loan	29-Jul-20	31-Dec-21
2017 ~ 2022	4-Apr-17	
Free	10-Aug-18	

----- Original Data -----			
Contract	Joined	Loan Date End	
2020 ~ 2025	29-Jan-20	NaN	
Jun 30, 2023 On Loan	9-Sep-20	30-Jun-23	
Free	1-Jan-11		
----- Resultant Data -----			
Contract	contractStartDate	contractEndDate	OnLoan
2020 ~ 2025	29-Jan-2020	29-Jan-2025	NO
2020 ~ 2023	9-Sep-2020	30-Jun-2023	YES
2011 ~ 9999	1-Jan-2011	31-Dec-9999	FREE

## Issues in the Dataset continued...

10. **Club** contains non ascii characters , numbers, fractions, Empty Values
11. **Column H** contains Invalid header (â†“OVA - should be OVER ALL RATING)
12. **Height** contains different measuring formats as feet and cms
13. **Weight** contains different measuring formats as KG and LBS

Club
VfL Osnabrück
VfL Wolfsburg
Viking FK
Viktoria Köln
Viktoria Plzeň
Villarreal CF
Vissel Kobe
Vitória Guimarães

H
â†“OVA
93
92
91
91
91
91

Height	Weight
194cm	100kg
5'11"	170lbs
5'10"	170lbs
196cm	120kg
191cm	97kg
186cm	100kg
188cm	100kg



# Problem Identification and Curation

## Identification :

1. Microsoft Excel Spreadsheet
2. Vizier visualization

## Curation :

1. Vizier Tool
2. Jupyter Notebook
3. Python Libraries :-
  - ⇒ **Pandas** - Pandas is a Python library for data analysis and data manipulation. It offers data structures and operations for manipulating numerical tables and time series.
  - ⇒ **Re** - Regex library for performing regular expression related operations over data
  - ⇒ **Datetime** - Date time library for curing various date formats and string to date conversions
  - ⇒ **Dateutil.relativedelta** - Utility functions for date operations like addition of years, etc.

# Python code Snippet

```
df['Height'] = df['Height'].str.replace("'",'.') #Converting to format
df['Height'] = df['Height'].str.replace(' ','K') #Converting to format

#Converting Feet to CM
def height(x):
    nu = float (re.findall('\d+', x )[0])
    strr = re.findall("[a-zA-Z]+", x)
    #print(strr)
    if 'K' in strr[0]:
        result = nu*30.48
        result = round(result)
        result = str(result) +'cm'
    else:
        result = x
    return result
```

## Curation of height Values

Converting height values from varied formats such as cms and feet-inches to cms only

Input	Output
175cm	175cm
5'10"	155cm

# Python code Snippet

```
def renameColumns(dict):  
    df.rename(columns=(dict), inplace=True)  
  
def getYearFrom(columnName, index):  
    date = df.iat[index, df.columns.get_loc(columnName)]  
    date = datetime.strptime(date, '%d-%b-%Y').date()  
    return date.year  
  
def valueAt(columnName, index):  
    return df.iat[index, df.columns.get_loc(columnName)]  
  
def modifyDateFormat(columnName, index):  
    date = valueAt(columnName, index)  
    date = datetime.strptime(date, '%d-%b-%y').date()  
    date = date.strftime('%d-%b-%Y')  
    df.iat[index, df.columns.get_loc(columnName)] = date
```

**renameColumns()** - To rename columns

**Example :** {'ID': 'EmpID'}

**getYearFrom()** - To retrieve an year value from date column in the dataframe object  
Expected Format of the Date column = '%d-%b-%Y'

**Example :** 29-Jan-2020 will give 2020

**modifyDateFormat()** - To modify date format from '%d-%b-%y' to '%d-%b-%Y'  
(Standardizing date format)

**Example :** 29-Jan-20 will become 29-Jan-2020

# Cleaned DataSet

ID	Name	LongName	photo	player	Nation	Age	Over A	POT	Club	Contract	Positic	Height	Weigh	Prefer	BOV	Best P	contractStartDate	contractEndDate	Value	Wage	Release	A/W	D/W	OnLoa
158023	L. Messi	Lionel Me	https://cd	http://sof	Argentina	33	93	93	2004 ~ 2021	RW, ST, CF	170cm	72kg	Left	93	RW	93	1-Jul-04	1-Jul-21	103000K	560K	138000K	Medium	Low	NO
20801	Cristiano C. Ronald	https://cd	http://sof	Portugal	35	92	92	2018 ~ 2022	ST, LW	187cm	83kg	Right	92	ST	92	10-Jul-18	10-Jul-22	63000K	220K	75000K	High	Low	NO	
200389	J. Oblak	Jan Oblak	https://cd	http://sof	Slovenia	27	91	93	2014 ~ 2023	GK	188cm	87kg	Right	91	GK	91	16-Jul-14	16-Jul-23	120000K	125K	159000K	Medium	Medium	NO
192985	K. De Bruy	Kevin De B	https://cd	http://sof	Belgium	29	91	91	2015 ~ 2023	CAM, CM	181cm	70kg	Right	91	CAM	91	30-Aug-15	30-Aug-23	129000K	370K	161000K	High	High	NO
190871	Neymar Jr	Neymar d	https://cd	http://sof	Brazil	28	91	91	2017 ~ 2022	LW, CAM	175cm	68kg	Right	91	LW	91	3-Aug-17	3-Aug-22	132000K	270K	166000K	High	Medium	NO
188545	R. Lewand	Robert Lew	https://cd	http://sof	Poland	31	91	91	2014 ~ 2023	ST	184cm	80kg	Right	91	ST	91	1-Jul-14	1-Jul-23	111000K	240K	132000K	High	Medium	NO
209331	M. Salah	Mohamec	https://cd	http://sof	Egypt	28	90	90	2017 ~ 2023	RW	175cm	71kg	Left	90	RW	90	1-Jul-17	1-Jul-23	120000K	250K	144000K	High	Medium	NO
212831	Alisson	Alisson R	https://cd	http://sof	Brazil	27	90	91	2018 ~ 2024	GK	191cm	91kg	Right	90	GK	90	19-Jul-18	19-Jul-24	102000K	160K	120000K	Medium	Medium	NO
231747	K. Mbapp	Kylian Mk	https://cd	http://sof	France	21	90	95	2018 ~ 2022	ST, LW, RW	178cm	73kg	Right	91	ST	91	1-Jul-18	1-Jul-22	185000K	160K	203000K	High	Low	NO
192448	M. ter Stej	Marc-And	https://cd	http://sof	Germany	28	90	93	2014 ~ 2022	GK	187cm	85kg	Right	90	GK	90	1-Jul-14	1-Jul-22	110000K	260K	147000K	Medium	Medium	NO
203376	V. van Dij	Virgil van	https://cd	http://sof	Netherlan	28	90	91	2018 ~ 2023	CB	193cm	92kg	Right	90	CB	90	1-Jan-18	1-Jan-23	113000K	210K	145000K	Medium	Medium	NO
208722	S. Manã	Sadio Ma	https://cd	http://sof	Senegal	28	90	90	2016 ~ 2023	LW	175cm	69kg	Right	90	LW	90	1-Jul-16	1-Jul-23	120000K	250K	144000K	High	Medium	NO
200145	Casemiro	Carlos He	https://cd	http://sof	Brazil	28	89	89	2013 ~ 2023	CDM	185cm	84kg	Right	89	CDM	89	11-Jul-13	11-Jul-23	90000K	310K	122000K	High	High	NO
192119	T. Courtoi	Thibaut C	https://cd	http://sof	Belgium	28	89	90	2018 ~ 2024	GK	199cm	96kg	Left	89	GK	89	9-Aug-18	9-Aug-24	82000K	250K	119000K	Medium	Medium	NO
167495	M. Neuer	Manuel N	https://cd	http://sof	Germany	34	89	89	2011 ~ 2023	GK	193cm	92kg	Right	89	GK	89	1-Jul-11	1-Jul-23	17000K	130K	47000K	Medium	Medium	NO
165153	K. Benzem	Karim Ber	https://cd	http://sof	France	32	89	89	2009 ~ 2022	CF, ST	185cm	81kg	Right	89	CF	89	9-Jul-09	9-Jul-22	83000K	350K	108000K	Medium	Low	NO
155862	Sergio Rai	Sergio Rai	https://cd	http://sof	Spain	34	89	89	2005 ~ 2021	CB	184cm	82kg	Right	89	CB	89	1-Aug-05	1-Aug-21	33000K	300K	50000K	High	Medium	NO
153079	S. Agñieri	Sergio Agñ	https://cd	http://sof	Argentina	32	89	89	2011 ~ 2021	ST	173cm	70kg	Right	89	ST	89	28-Jul-11	28-Jul-21	83000K	300K	98000K	High	Medium	NO
202652	R. Sterling	Raheem S	https://cd	http://sof	England	25	88	90	2015 ~ 2023	LW, RW	170cm	69kg	Right	88	LW	88	14-Jul-15	14-Jul-23	114000K	270K	139000K	High	Medium	NO
215914	N. Kantã	N'Golo Ka	https://cd	http://sof	France	29	88	88	2016 ~ 2023	CDM, CM	168cm	70kg	Right	88	CDM	88	16-Jul-16	16-Jul-23	78000K	190K	96000K	Medium	High	NO
212622	J. Kimmic	Joshua Ki	https://cd	http://sof	Germany	25	88	90	2015 ~ 2023	CDM, RB	176cm	73kg	Right	88	CDM	88	1-Jul-15	1-Jul-23	103000K	145K	112000K	High	Medium	NO
211110	P. Dybala	Paulo Dyk	https://cd	http://sof	Argentina	26	88	89	2015 ~ 2022	CF, CAM	177cm	75kg	Left	89	CAM	89	1-Jul-15	1-Jul-22	109000K	190K	122000K	Medium	Medium	NO
210257	Ederson	Ederson S	https://cd	http://sof	Brazil	26	88	91	2017 ~ 2024	GK	188cm	86kg	Left	88	GK	88	1-Jul-17	1-Jul-24	92000K	195K	103000K	Medium	Medium	NO
202126	H. Kane	Harry Ka	https://cd	http://sof	England	26	88	89	2010 ~ 2024	ST	188cm	89kg	Right	88	ST	88	1-Jul-10	1-Jul-24	190000K	220K	140000K	High	High	NO
162835	S. Handar	Samir Ha	https://cd	http://sof	Slovenia	35	88	88	2012 ~ 2021	GK	193cm	92kg	Right	88	GK	88	1-Jul-12	1-Jul-21	10000K	100K	27000K	Medium	Medium	NO
201024	K. Koulibz	Kalidou K	https://cd	http://sof	Senegal	29	88	88	2014 ~ 2023	CB	187cm	89kg	Right	88	CB	88	1-Jul-14	1-Jul-23	76000K	140K	85000K	Medium	High	NO
183277	E. Hazard	Eden Haz	https://cd	http://sof	Belgium	29	88	88	2019 ~ 2024	LW, ST	175cm	74kg	Right	88	LW	88	1-Jul-19	1-Jul-24	89000K	350K	118000K	High	Medium	NO
182521	T. Kroos	Toni Kroo	https://cd	http://sof	Germany	30	88	88	2014 ~ 2023	CM	183cm	76kg	Right	88	CM	88	17-Jul-14	17-Jul-23	87000K	310K	112000K	Medium	Medium	NO
194765	A. Griezmi	Antoine G	https://cd	http://sof	France	29	87	87	2019 ~ 2024	ST, CF, LW	176cm	73kg	Left	87	ST	87	12-Jul-19	12-Jul-24	79000K	290K	103000K	Medium	Medium	NO
233049	J. Sancho	Jadon Sa	https://cd	http://sof	England	20	87	93	2017 ~ 2023	RM, LM, C	180cm	76kg	Right	89	CAM	89	31-Aug-17	31-Aug-23	124000K	82K	132000K	High	Medium	NO
231281	T. Alexanc	Trent Alex	https://cd	http://sof	England	21	87	92	2015 ~ 2024	RB	180cm	69kg	Right	87	RB	87	7-Oct-15	7-Oct-24	114000K	110K	115000K	High	Medium	NO
218667	Bernardo	Bernardo	https://cd	http://sof	Portugal	25	87	88	2017 ~ 2025	RW, CAM	173cm	64kg	Left	87	RW	87	1-Jul-17	1-Jul-25	95000K	230K	115000K	High	High	NO
216267	A. Roberts	Andrew R	https://cd	http://sof	Scotland	26	87	89	2017 ~ 2024	LB	178cm	64kg	Left	87	LB	87	21-Jul-17	21-Jul-24	90000K	155K	102000K	High	High	NO
212218	A. Laporte	Aymeric L	https://cd	http://sof	France	26	87	90	2018 ~ 2024	CB	189cm	85kg	Left	88	CB	88	30-Jan-18	30-Jan-24	92000K	200K	107000K	Medium	Medium	NO
212198	Bruno Fer	Bruno Mi	https://cd	http://sof	Portugal	25	87	90	2020 ~ 2025	CAM, CM	179cm	69kg	Right	88	CAM	88	30-Jan-20	30-Jan-25	105000K	195K	124000K	High	High	NO
209499	Fabinho	Fãbio He	https://cd	http://sof	Brazil	26	87	89	2018 ~ 2023	CDM	188cm	78kg	Right	87	CDM	87	1-Jul-18	1-Jul-23	88000K	155K	104000K	Medium	High	NO
200104	H. Son	Heung Mi	https://cd	http://sof	Korea Rep	27	87	87	2015 ~ 2023	LM, CF	183cm	78kg	Right	87	LM	87	28-Aug-15	28-Aug-23	85000K	165K	99000K	High	High	NO
201942	Roberto F	Roberto F	https://cd	http://sof	Brazil	28	87	87	2015 ~ 2023	CF	181cm	76kg	Right	87	CF	87	6-Jul-15	6-Jul-23	81000K	190K	96000K	High	High	NO
193041	K. Navas	Keylor Na	https://cd	http://sof	Costa Rica	33	87	87	2019 ~ 2023	GK	185cm	80kg	Right	87	GK	87	2-Sep-19	2-Sep-23	26000K	110K	50000K	Medium	Medium	NO

# **Challenges and Future Scope**

## **Challenges Faced :**

1. Find unclean dataset.
2. Inter-attribute relationship identification.
3. Data representation Issues.
4. Vizier installation.

## **Future scope and Statistics :**

1. Calculate Data Nullability Rate
2. Calculate Missing values and Inconsistent values
3. Ready for Data analysis and modeling.
4. High-value prospects can be identified more easily.
5. Ready to be put under GDPR(Global Data Protection Regulation) Implementation (If required)



THANK YOU!

