M Gmail

Aditya Avs <avs.aditya@gmail.com>

INSAID Capstone Project Observation
11 messages

---

sudeep raj <sudeep48raj93@gmail.com>   Fri, Aug 2, 2019 at 2:11 PM
To: abhishekashyap21@gmail.com, Alok.k.mallik@gmail.com, sudhir.takke@gmail.com, bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com, roumildhiman@gmail.com, avs.aditya@gmail.com, feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com, paresh.404@gmail.com, raasik1994@outlook.com, liza.mohanty01@gmail.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Hi All,

We have to submit our observations and finding by today 2nd August 2019.

Please sync-up on Slack channel.
https://app.slack.com/client/TLULX21LM/CLWG6GQJ3

Regards
Sudeep

Sudhir Takke <sudhir.takke@gmail.com>   Fri, Aug 2, 2019 at 2:14 PM
To: sudeep raj <sudeep48raj93@gmail.com>
Cc: abhishekashyap21@gmail.com, Alok.k.mallik@gmail.com, bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com, roumildhiman@gmail.com, avs.aditya@gmail.com, feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com, paresh.404@gmail.com, raasik1994@outlook.com, liza.mohanty01@gmail.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Did you manage to connect to sql-connector thru python notebook
I am unable to install sql-connector package
Let me know if anyone alredy did this and uploaded data
[Quoted text hidden]

Aditya Avs <avs.aditya@gmail.com>   Fri, Aug 2, 2019 at 2:27 PM
To: Sudhir Takke <sudhir.takke@gmail.com>
Cc: sudeep raj <sudeep48raj93@gmail.com>, abhishekashyap21@gmail.com, Alok.k.mallik@gmail.com, bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com, roumildhiman@gmail.com, feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com, paresh.404@gmail.com, raasik1994@outlook.com, liza.mohanty01@gmail.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Hi Sudhir,

PFA the code for the same.
For installation, please use pip install mysql.connector

@Team Request you all to be please be more active because this is a group activity and we need to work together as a team.
We would gain more insights when we work together as a team.
Request all of you to spend some time today and make our project successful.

Special thanks to Sudeep for co-ordinating the same.

Thanks and Regards
Aditya
[Quoted text hidden]
--

A.V.S.Aditya

 Knowledge is the antidote to fear.
Ralph Waldo Emerson

📄 **importData.py**
2K

---

sudeep raj <sudeep48raj93@gmail.com>   Sat, Aug 3, 2019 at 9:58 PM
To: Aditya Avs <avs.aditya@gmail.com>
Cc: Sudhir Takke <sudhir.takke@gmail.com>, abhishekashyap21@gmail.com, Alok.k.mallik@gmail.com,
bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com, roumildhiman@gmail.com,
feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com, paresh.404@gmail.com,
raasik1994@outlook.com, liza.mohanty01@gmail.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Hi Everyone,

Please share your finding over the mail / slack, so that we can update our findings on the sheet share by INSAID.

Regards
Sudeep Raj
[Quoted text hidden]

Liza Mohanty <liza.mohanty01@gmail.com>   Sat, Aug 3, 2019 at 10:38 PM
To: sudeep raj <sudeep48raj93@gmail.com>
Cc: Aditya Avs <avs.aditya@gmail.com>, Sudhir Takke <sudhir.takke@gmail.com>, abhishekashyap21@gmail.com,
Alok.k.mallik@gmail.com, bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com,
roumildhiman@gmail.com, feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com,
paresh.404@gmail.com, raasik1994@outlook.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Hi All,

As of now could come up with the following findings/analysis.Please let me know if you have any comments. Am still
working on the eda part and feature engineering.Once done will share my working file.

Employees Dataset details:

There are 3 datasets related to the employees:

1. department records. There are 11 records.`'dept_id', 'dept_name', 'dept_head'`

2. employee_details_data : There are 14245 records of employees. `employee_id', 'age', 'gender', 'marital_status'`

3. employee_data :There are 14150 records of employees. Features included are `avg_monthly_hrs, department,filed_complaint, 'last_evaluation', 'n_projects', 'recently_promoted' , 'salary', 'satisfaction', 'status','tenure', 'employee_id'`

```
Challenges : there are records with missing and incorrect values. There are some
duplicate records as well.Also need to join other two datasets with employee
dataset to build a good model for predicting the status (employed or left).
```

1. Missing Values

Only employee dataset has missing values. Following features or columns have missing values:

```
department (707 records),
```

```
filed_complaint (12104 records),
last_evaluation (1487 records),
recently_promoted (13853),
satisfaction (150 records),
tenure (150 records)
Missing values in percentage :
recently_promoted     97.90
filed_complaint       85.54
last_evaluation       10.51
department             5.00
tenure                 1.06
satisfaction           1.06
```

Drop the features/ columns - recently_promoted and filed_complaint as the missing percentage is more than 80% and they wont add any value for further analysis.

Drop records with missing department , tenure and satisfaction ,last evaluation as the percentage of records/rows with the missing values is very less.So it's safe to remove them instead of trying to find the replaceable values (mean/mode/median).

2. Update the correct department information in the employee dataset

Few employee records with department as '-IT' replace with correct department id i.e. D00-IT from the department dataset

3. Duplicate values

Only employee dataset has duplicate records including records with employee id as 0.

There are 4 records with employee id as 0 . Delete these records.

Then there are 24 duplicate records. Delete the 24 duplicate records.

4.Update the employee dataset before building the right model

The following features from the employee details dataset needs to be added to the employee dataset gender , marital_status to the employee dataset.

5.Encoding the categorical features :

Change categorical features in emp details table gender , marital_status,salary,department, status

Numerical features : employee id, age,tenure,satisfaction,n projects, last_evaluation,salary.

Working on data analysis through visualizations and feature engineering.

Thanks,

Liza
[Quoted text hidden]

Sudhir Takke <sudhir.takke@gmail.com>   Sat, Aug 3, 2019 at 11:16 PM
To: Aditya Avs <avs.aditya@gmail.com>

Thanks Aditya
Guys - I may be be late in responding/contributing... but I will be onto it

Thanks & Regards,

Sudhir Takke
+919960657700

I'm protected online with Avast Free Antivirus. Get it here — it's free forever.

[Quoted text hidden]

Aditya Avs <avs.aditya@gmail.com>   Mon, Aug 5, 2019 at 2:18 AM
To: Sudhir Takke <sudhir.takke@gmail.com>

Hi Sudhir,

You just replied to me.
I suggest you please start using the slack channel for communication.

Thanks and Regards
Aditya
[Quoted text hidden]

Aditya Avs <avs.aditya@gmail.com>   Mon, Aug 5, 2019 at 2:33 AM
To: Liza Mohanty <liza.mohanty01@gmail.com>
Cc: sudeep raj <sudeep48raj93@gmail.com>, Sudhir Takke <sudhir.takke@gmail.com>, abhishekashyap21@gmail.com,
Alok.k.mallik@gmail.com, bhattdrupad2016@gmail.com, ksreddy1141@gmail.com, bhuyanpk@gmail.com,
roumildhiman@gmail.com, feerojkumarp@gmail.com, avisesudhakarrao@gmail.com, ashishghadge@gmail.com,
paresh.404@gmail.com, raasik1994@outlook.com, sushilbshinde@gmail.com, sibrajb@gmail.com

Hi Liza,

Excellent observations.

Me and Sudeep had a zoom call on Friday.
We also discussed few other observations.

I have done some EDA.
Created two notebooks one with replacing the data and other deleting the same.
I have them in my github.
https://github.com/avsaditya/INSAID-GCD/tree/master/CapstoneProject

Sorry for not being able to organize a call on Sunday.
Also, I think google hangouts seems a better option for calls rather than zoom.

Let us try to fix something up for Monday night.
Also, please share your EDA notebook.
We can put together the best things into a single sheet which would be suitable for team submission.

@Team - Please come up with suggestions and your observations.
Please start by joining the slack channel.
If you are unable to join the slack, please reply to the email chain.

Thanks and Regards
Aditya
[Quoted text hidden]

Aditya Avs <avs.aditya@gmail.com>   Fri, Aug 9, 2019 at 12:46 AM

To: Liza Mohanty <liza.mohanty01@gmail.com>, sudeep raj <sudeep48raj93@gmail.com>, Sudhir Takke
<sudhir.takke@gmail.com>, abhishekashyap21@gmail.com, Alok.k.mallik@gmail.com, bhattdrupad2016@gmail.com,
ksreddy1141@gmail.com, bhuyanpk@gmail.com, roumildhiman@gmail.com, feerojkumarp@gmail.com,
avisesudhakarrao@gmail.com, ashishghadge@gmail.com, paresh.404@gmail.com, raasik1994@outlook.com,
sushilbshinde@gmail.com, sibrajb@gmail.com

Hi Team,

Please reply back with your response.

It is been 2 weeks since the start of the project and most of you have not even responded to the email thread nor in
the slack group.
This is not the way we would work like a team.
The idea of forming this team is to make effective use of collective knowledge.
It would give different insights to each of us and widen our horizon.

I request you, for the last time, to start contributing actively.
Please start by responding to the mail or in slack.

Thanks and Regards
Aditya
[Quoted text hidden]

Raasik Ravindran <raasik1994@outlook.com>   Fri, Aug 9, 2019 at 12:31 PM
To: Aditya Avs <avs.aditya@gmail.com>, Liza Mohanty <liza.mohanty01@gmail.com>, sudeep raj
<sudeep48raj93@gmail.com>, Sudhir Takke <sudhir.takke@gmail.com>, "abhishekashyap21@gmail.com"
<abhishekashyap21@gmail.com>, "Alok.k.mallik@gmail.com" <Alok.k.mallik@gmail.com>,
"bhattdrupad2016@gmail.com" <bhattdrupad2016@gmail.com>, "ksreddy1141@gmail.com" <ksreddy1141@gmail.com>,
"bhuyanpk@gmail.com" <bhuyanpk@gmail.com>, "roumildhiman@gmail.com" <roumildhiman@gmail.com>,
"feerojkumarp@gmail.com" <feerojkumarp@gmail.com>, "avisesudhakarrao@gmail.com"
<avisesudhakarrao@gmail.com>, "ashishghadge@gmail.com" <ashishghadge@gmail.com>, "paresh.404@gmail.com"
<paresh.404@gmail.com>, "sushilbshinde@gmail.com" <sushilbshinde@gmail.com>, "sibrajb@gmail.com"
<sibrajb@gmail.com>

Hi Aditya,

Sorry was kind of busy with my office works for the last few weeks. I have already joined Slack and i have
started doing EDA on the data. I Will try to share my analysis EOD.

I have been following the Email and slack notifications and will be able to participate actively hereafter.

Regards
Raasik.R

From: Aditya Avs <avs.aditya@gmail.com>
Sent: Friday, August 9, 2019 12:46 AM
To: Liza Mohanty <liza.mohanty01@gmail.com>; sudeep raj <sudeep48raj93@gmail.com>; Sudhir Takke
<sudhir.takke@gmail.com>; abhishekashyap21@gmail.com <abhishekashyap21@gmail.com>;
Alok.k.mallik@gmail.com <Alok.k.mallik@gmail.com>; bhattdrupad2016@gmail.com
<bhattdrupad2016@gmail.com>; ksreddy1141@gmail.com <ksreddy1141@gmail.com>;
bhuyanpk@gmail.com <bhuyanpk@gmail.com>; roumildhiman@gmail.com <roumildhiman@gmail.com>;
feerojkumarp@gmail.com <feerojkumarp@gmail.com>; avisesudhakarrao@gmail.com
<avisesudhakarrao@gmail.com>; ashishghadge@gmail.com <ashishghadge@gmail.com>;
paresh.404@gmail.com <paresh.404@gmail.com>; raasik1994@outlook.com
<raasik1994@outlook.com>; sushilbshinde@gmail.com <sushilbshinde@gmail.com>; sibrajb@gmail.com
<sibrajb@gmail.com>

Subject: Re: INSAID Capstone Project Observation

[Quoted text hidden]

Sibraj Banerjee <sibrajb@gmail.com>   Fri, Aug 9, 2019 at 12:57 PM
To: Aditya Avs <avs.aditya@gmail.com>

Hi Aditya,
Due to illness I have been unable to respond back on Slack and in this group. Sorry for delayed response. I am looking into the data and will update you on the same.

Thanks & Regards,
Sibraj Banerjee.

[Quoted text hidden]