

Summer Internship Report

Submitted

by

Kreethi Mishra – AP19110010424




Department of Computer Science and Engineering

SRM University-AP, Andhra Pradesh, India

July - 2021

DATA SHEET

Roll Number	:	AP19110010424
Name of the student	:	Kreethi Kumari Mishra
Branch & Section	:	CSE C
Batch	:	2019-2023
Type of internship	:	Industrial - Training & Internship
Company Name/Institute Name	:	ShapeAI
Company/Institute Website	:	https://www.shapeai.tech/
Start Date (MM/DD/YYYY)	:	06/01/2021
End Date (MM/DD/YYYY)	:	08/10/2021
Duration (No. of days)	:	70
Status of the internship	:	Ongoing
Name of internship mentor (SRM Faculty)	:	Dr. Md. Muzakkir Hussain

Signature of the student	:	
--------------------------	---	--

ACKNOWLEDGEMENT

I would like to express my special thanks to my mentor for this Industrial Training and internship is Mr. Madhav Jauhari. He gave me this wonderful opportunity to do an internship under him. The way he taught helped me to grab everything to complete the project.

I would also like to thank SRM University AP for accepting our Industrial Internship. This internship has helped me to improve my skills in a better way and also helps in my future studies.

Finally, an honourable mention goes to my family and friends for their understanding and support of me in completing this project. Without the help of the people mentioned above, I would have faced many difficulties while doing this.

Kreethi Mishra
AP19110010424

TABLE OF CONTENTS

Sl. No.	Content	Page No.
1	Introduction	5
2	Objective of the Internship	5
3	Skills acquired through the internship	5
4	Overview of the project carried out during internship	5
5	Results	6
6	Conclusion	8

I. INTRODUCTION

A data analyst is someone who scrutinises information using data analysis tools. The meaningful results they pull from the raw data help their employers or clients make important decisions by identifying various facts and trends. Typical duties include: using advanced computerised models to extract the data needed.

Eg: Amazon Fresh and Whole Foods is a perfect example of how big data can help improve innovation and product development. Amazon leverages big data analytics to move into a large market. Focusing on big data analytics, Amazon whole foods is able to understand how customers buy groceries and how suppliers interact with the grocer. This data gives insights whenever there is need to implement further changes.

II. OBJECTIVE OF THE INTERNSHIP

The objective of this internship is to give the student an opportunity to reflect on the professional aspects of the internship experience and the skills that were learned. It helped to learn new skills and great experience through it. It helped us to get a grip on python language. we have also learned about data visualization, machine learning algorithms and their importance.

III. SKILLS ACQUIRED THROUGH THE INTERNSHIP

- Python
- Problem solving
- Data visualization
- Machine learning

IV. OVERVIEW OF THE PROJECT/WORK CARRIED OUT DURING INTERNSHIP

I worked on a project named salary Prediction. There is a data set given which contains details of Name, Job Titles, Agency Id's, Hire date, Annual salary and gross Pay this data is of the employees of Baltimore. We need to predict their salaries. And we also need to analyse the given data by dividing the data and we find out the relations of the data by comparing salary and hire month like which months the companies are hiring the more employees by using bar graphs and correlations between the data. I also divided the data set into parts and find whether it is good model.

V. RESULTS/OUTPUT

Head of the csv:-

<>

📁

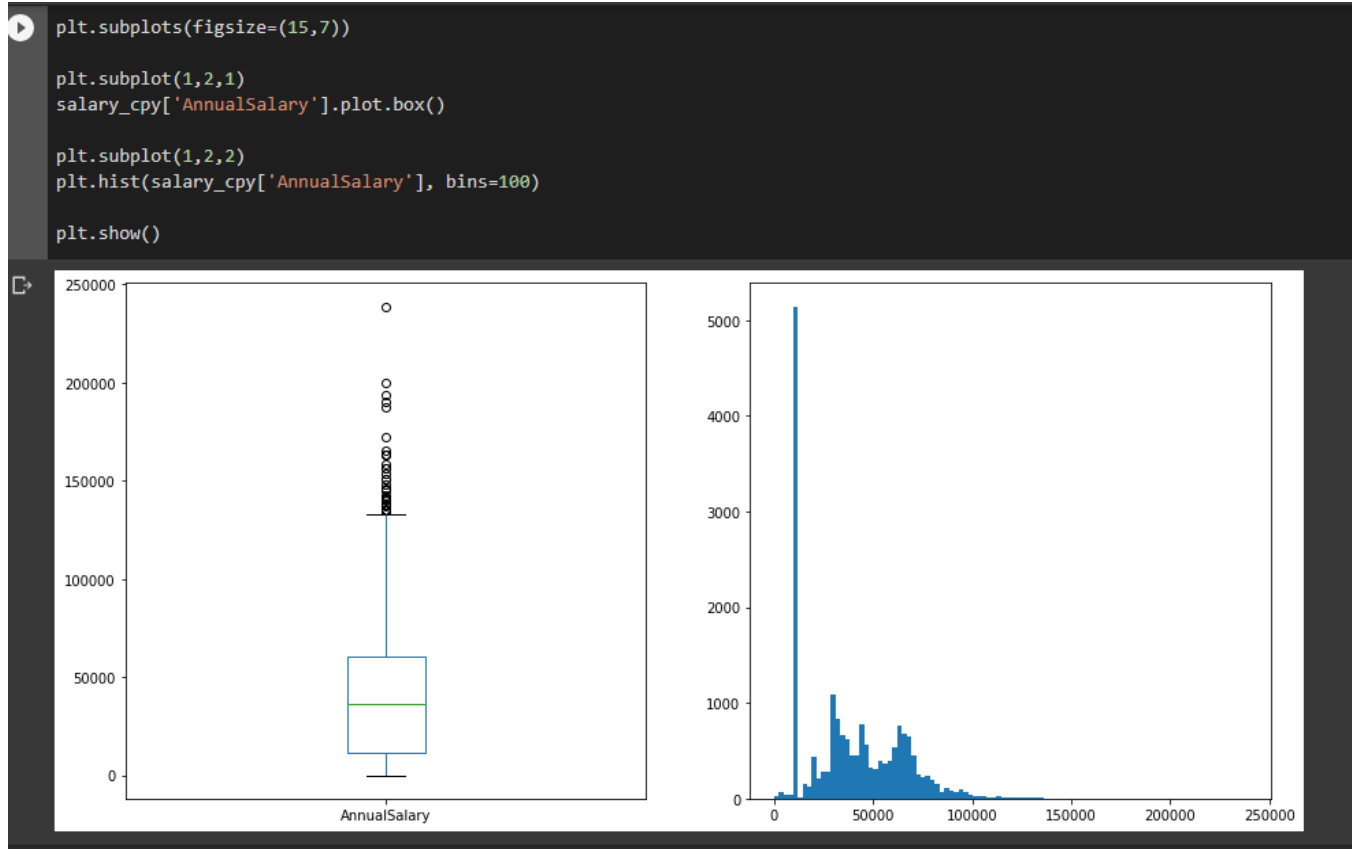
▶

```
# Import the data set
salary_org = pd.read_csv("/content/train.csv")

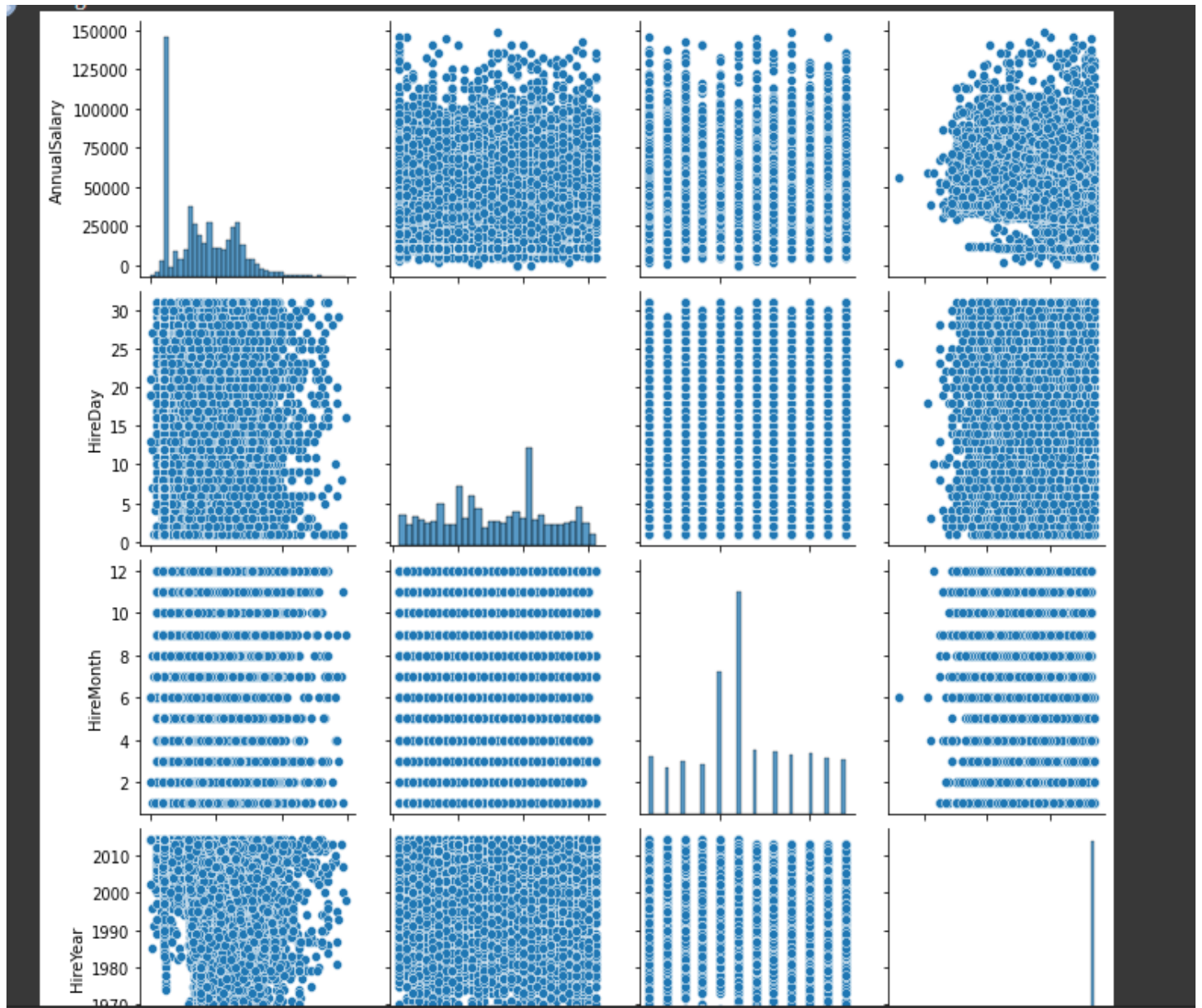
# Print top 5 rows of dataset
salary_org.head()
```

	Name	JobTitle	AgencyID	Agency	HireDate	AnnualSalary	GrossPay
0	Aaron,Keontae E	AIDE BLUE CHIP	W02200	Youth Summer	06/10/2013	\$11310.00	\$873.63
1	Aaron,Patricia G	Facilities/Office Services II	A03031	OED-Employment Dev	10/24/1979	\$53428.00	\$52868.38
2	Aaron,Petra L	ASSISTANT STATE'S ATTORNEY	A29005	States Attorneys Office	09/25/2006	\$68300.00	\$67439.19
3	Abaineh,Yohannes T	EPIDEMIOLOGIST	A65026	HLTH-Health Department	07/23/2009	\$62000.00	\$58654.74
4	Abbene,Anthony M	POLICE OFFICER TRAINEE	A99416	Police Department	07/24/2013	\$43999.00	\$39686.95

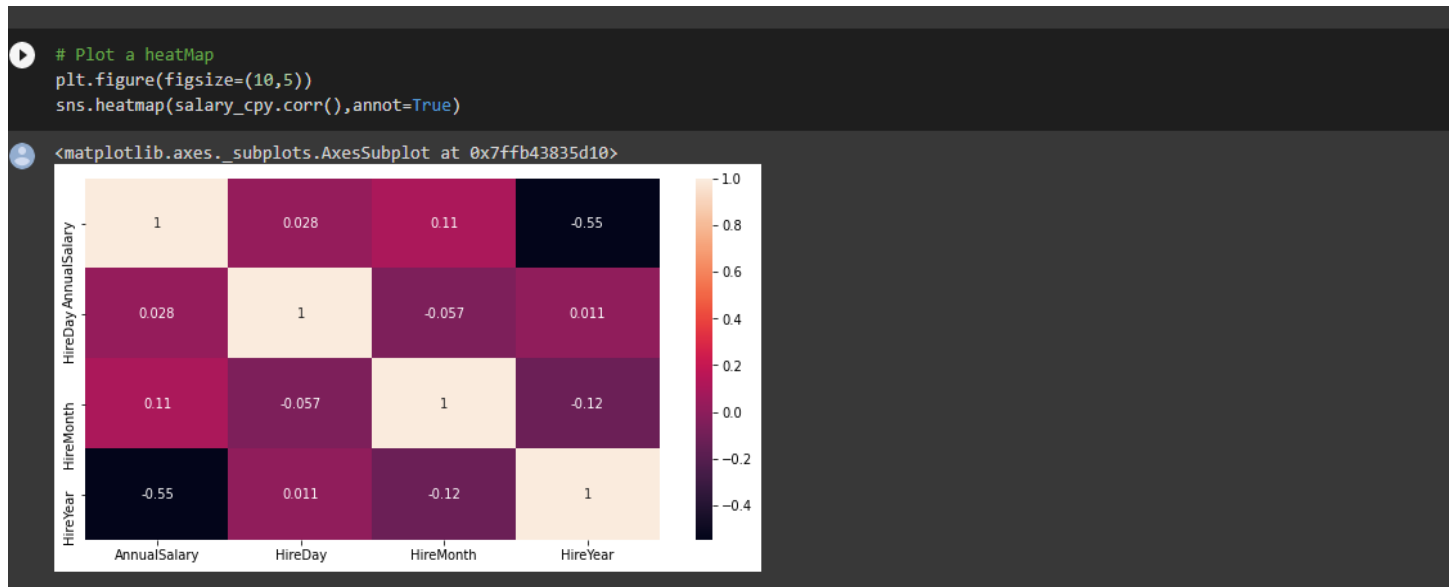
Box Plot:-



Pair Plot:-



Correlation:-



Coefficient and intercept:-

```
# Print coef
print("Coef are:",salary_reg.coef_)

#print intercept
print("Intercept is",salary_reg.intercept_)
```

Coef are: [22395.65357851 2798.67752824 -1694.55993659 -22.59156434
-62.65873076 -2049.21601256]
Intercept is 39628.221303493134

VI. CONCLUSION

- From the above project I concluded that the agency Id, JobTitle and hire year are the major columns for explaining the variance of the annual salary which has the highest coefficient.
- The Hire year is negatively related to the annual salary they both have almost opposite intercepts.