



Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	740683
Project Title	Doctors Annual Salary prediction
Maximum Marks	6 Marks

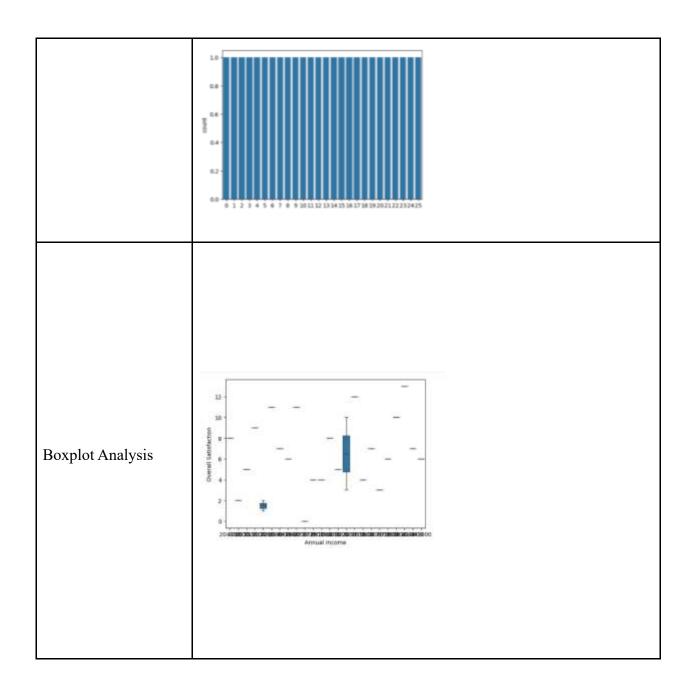
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Desc	riptio	n							
Data Overview	Dimension: 28 rows × 8 columns Descriptive statistics:									
	Ð	Specialty	Arrical Discuss	real retrly Compensated	Overall satisfaction	Setzefied Drome	inuld Chance Pedicine Again	smuld Cheese the	Survey Respondents By Specialty	
	sount	26 200000	28.000000	26.000000	26.100000	28.000000	38.000000	26,000000	36,000000	
	present	12.800000	201423-010023	5.103846	& SKITTERED.	5.269231	8.823077	9.853649	2.200251	
	100	YARRED	71044.872281	5.120168	3.516741	3.317301	5.030404	5.695162	2.794890	
	resize.	6.000000	204000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
	25%	R 280000	228000.000000	2,290000	4.000000	2.250000	5.290000	6.250000	0.000000	
	58%	12.503100	201500.000000	5.000300	6.069000	5.700000	14.900000	8.500000	1.000000	
	29%	18,750000	398750.000000	7.790000	8.750000	3.790000	16,0000081	19,790000	3.790000	
	mee	25.000000	A0000.00000	11,000000	13,000000	12.000000	10,000001	20.000000	8,000000	
	500000				11133333	New Control		35,173,0035	17,0000	
Countplot Analysis										

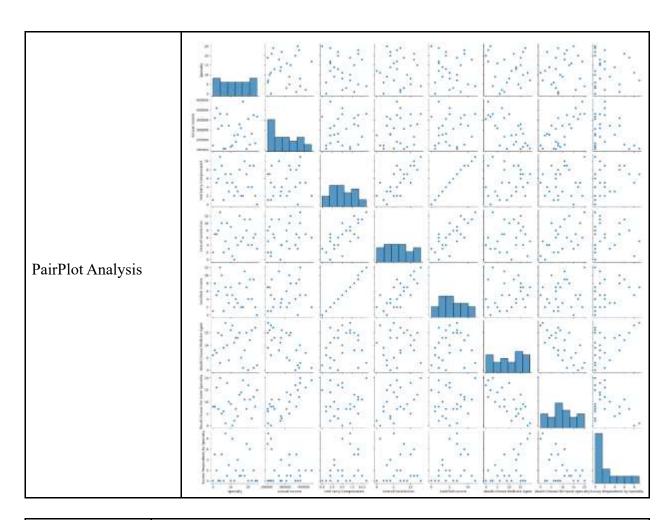












Outliers and			
Anomalies	-		

Data Preprocessing Code Screenshots





Loading Data	25	Aperciality	Securit Secure	Feel Febrily Compensated	Overal3 Settiofection	SHESHESHE SHIDNE	model theres resticine Again	mould (house the fame Specialty	turney texpositetts by specialty	8
		Ottopesko	4000	2.44	288	2.44	148	0.86	0.00	
	. 4	Cartings	410000	0.49	554	0.46	0.00	9.67	0.00	7
	2	Demokrapy	36100	0.00	0.00	0.86	11.60	6.74	101	
	3	Owitoermoney	36000	0.46	3.97	5.49	3.61	0.60	5.65	
	4	Ratings	375000	9.38	151	1.00	5.40	0.53	9.00	
		Shirings	367(6)	0.42	1.00	0.42	8.01	0.00	8.01	
Louding Data		Anestreetings	-	3.86	134	8.50	120	0.40	106	
	7	Plants Surgery	20000	5.47	0.91	0.47	3,47	0.34	0.01	
		Oroman	309000	0.00	5.00	0.00	0.00	654	100	
	,	Streepeny Neitone	322999	0.86	9.97	0.60	0.00	0.44	0.00	
	16	Correct Surgery	NAME	0.40	100	0.40	0.04	681	104	
	. 10	Cyristiamology	30000	0.44	116	0.44	0.00	9.30	140	
	12	Ottor Gen	30000	6.00	136	9.90	3.00	5.40	681	
	11	Pulmmany Medicine	281000	5.47	1.01	8.47	149	637	541	
Data Transformation	<pre>s, brade = s, trad [PELINE] (bry [N], ", regree (bra) = attpac (finat') / IRE s, best = s, trad [region([N,1]", ", regree (bra) = attpac (finat') / IRE y trade = y trade = applica([N,1]", ", regree (bra) = attpac (finat') / IRE y_trade = y trade (applica([N,1]", ", regree (bra) = attpac (finat') / IRE impoter s = timple Departer (attractors (application) = attpace (finat') / IRE impoter s = timple Departer (attractors (application) = attractors (application) = attractors</pre>									
Feature Engineering	Attacl	ned the co	odes i	n final sı	ıbmissic	on.				
Save Processed	_									