	Name: Chhagan Kumavat Pollno: 20230201067 Assignment 3
	What is duta preprocessing?  Data preprocessing:  Data preprocessing:  Data preprocessing:  The main step of data preprocessing include.  Data cleaning:  Removing or correcting incounted data and daling with missing value and eliminating moise
2.	Data Integration! Combining data from different sources into a cornerent dataset.  Data Bedunction!
fr <sub>(m</sub> 4)	Reducing the volume of data by aggregating, summerizing or simply compares into form.  Data Transformation:  Normalizing or scaling data to a countistent formate.
· · · · ·	The goal of data preprocessing is to improve the quality of the data and make suitable for analysis or modeling.
	alpha lauta

	Newton many and a second way	* assues of Data Antegration:
	Data	) Schema integration of sheet matching Rodan
adle or	Cleaning	danty
3.71.5		(2) Reduction of deeduction of data values
- st		3) No change vor difficult to change
	Rata Data	of the detailing to the detail
	Rata Data Transformation preprocessing Integration	* The steps involved in FTL an:
	All some and a some a some and a some a some and a some and a some	- Extract: Retrieving raw data from different
	Charles Charles Charles	Sources vivonica i utilous and made and
d daline	Data	Sources System such as I database Apis or Mat files.
3	Reduction	otal so tilevo
		- Gransformer:
	a Oat Tatearntie :	cleaning transforming a empicing the data
nakai	Data proprocessing.	cleaning transforming a empiching the data
	the state of the s	militarget is system who ai hour starily
		This include data cleaning mormalization,
02.	What is ETI in the context of Data Integration	aggregation dapplying business rules.
)	ETL: - when a gooder with minutes the	
	ETL stand for Extract Transform load and it	Systems unich could be a data into target data mart or databaic.
	15 a process used in data integration to	Systems unich could be a data inchouse,
	more data from various sourcess to a target	data mart or database.
51	data Warehouse.	
	- torrot	- FTL:-
- 11	Lata Integration:	disparate sources enaing data quality and
only y	Difference type or data from diffrent	dispersate sources enjoying data quality and
7	Difference type of data from diffrent Sourcess conherent together.	molariza data for analysis.
~~~	A CALL DAY OF THE STATE OF THE	- Language policy personal -
	Clypes	note introduced in some principality
	1) Ctype coopling - Actual data Cones together. 2) lusse Copling - using interface but not	charing restances the performance of muchine.
	2) lusse Copling - using interface but not	leaving nodels and other analytical techniques.
	actual data	Veril 1 de la constante de la

DATE //

and it is a short during dat	
Q3. Why is Rata-cleaning important during data	- Enhance Business Decision - making -
preprocessing?  Data cleaning:	Reliable Jato trinsignts, made possible by
Data cleaning:	data cleaning isupport informed builness.
lote leanly literated	decision and strategy developments
because it ensures the according commency	F
and religibility of the data.	,
and reliability of the data.	CP4. What are some common Data Transformation
- Clean data: is caucial for several masons	techniques wed in 167171
· Improves Data Quality: Removing errors,	3 Data Grans: formation in ETL involves converting
inconsistencies and duplicates enhances the overal	data form its original bornat to a bornat
quality of data.	that is quitable for analysis to reporting.
quanti of data	that is suitable for analysis & reporting.
Enhance analysis Accuracy:	
clean data dead to mor accurate 1	- Data Normalization  Scaling data to a standard range uch
reliable result in data analysis - 4 machine	(0, 1) or [=), 1) to ensure constitute.
élecinina madalication	Co) 15 of Company
cloaming modelines en suite de	- Data Aggregation:
- Reduce Moise:	Summaning do to a higher devel,
Eliminating implement or includant information	Sign of colonial culture that we are right of
helps in a forcing on med tils and a Rollems cond	Such as calculating took is average or
in sights. I would be took or or	sil coon the source such as the course
1119191113	- Pata Cardina is
Increases Efficiency :-	- Rata Encoding:- (onweighing data with additional information,
Increases Efficiency:- Clean data reduce! the time and effort	Such as adding jeolocation data based on
required to process of analyze the data.	TP addresses
Melowo (St. stella strata)	showed for more it bose of the contractor with
- Improve Model performance:	a contraction of the same of the
	Removing Moderels data
By remains errors and inconsistencies data	michaela
cleaning renhances the performance of machine.  learning models and other analytical techniquee.	- Duly Conting '
analytical techniquee.	- Data Sarting:  And Arranging data in particular order.
	Hay Fill dryling water 181

Printer to a	thouse the policy as the things were
/ DATE / /	60 110 So 2 SO FIGE NO. 11 Q
	10 to the first
- I have the survey of the	
- Data Cleaning!	Data doading in the second is
Removing unwanted data or records hardas	This is final step when re transformat
on Specific conditions. Internal	close if mared into target system such as a
on specific conditions littering lilling In missing values, and correcting errors to ensure day	data warehouse. The focus is an enjurasa
quality.	that the data is accurating and efficiency
	lood into the target system for E
Reshaping !- 12 some some some some some	
Converting data structures, such as pivoting tally	
Converting data structures, such as pivoting talls or aggregating data, to prepare it for analysis	Source Transform
hard love of microl former de a later of	data propare for
- porthagon it involves or statement it to the	Light data in a substantial control of the state of the s
0'5 How along Octo a loading differ from Data Extraction	migue Institution mont Staging of ma - 1 min
of Transformation in ET-Lile	, wishing
Data loading is the final step in the FI	
pocess of differ from data extraction of transforme	Determine of row
tion in following ways:	in in white his not reside lifer its umagate.
- Wate Hage pation .	in the bout pritainity alob laining to
Data Extraction of data singuistics	Net 1
This is the first step where now data is	2
collected from virtual source system. The	(6/8/94) Match changes valive
princing focus is on retrieving data virthout any	to any me of store con with texisting diament
prodification.	proposed with state of timeter value, our son
e witnessed in withhe office at the prixary and	15000 Louise incer
- Rota Transformation: 10000 10000 1000	
This step involve cleans transforming and enhancing	Le mastanai la villas de la
the extraction data to need the reput set remember	overside correcte a new puch dow change
et the towest systems It day goods - all me	attribute diamontion record value to old
pushess rules, normalizing Hata di gragorogation	getrabuter fields.
information.	
- 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1	Dota Cloading losse basel 2
- ishall selection of other property of	in beautiful dela gradal - Suches organia
	will belt transmit some made the this make
	Euros structegies.
	· Carlo and a second