UNIVERSITÀ DEGLI STUDI FIRENZE

Baker Hughes

HISTORICAL DATA ANALYSIS TO SUPPORT THE CLASSIFICATION OF TURBINE AND COMPRESSOR COMPONENTS AND PREDICT FUTURE DEMAND

Canditate: MATTEO D'SOUZA

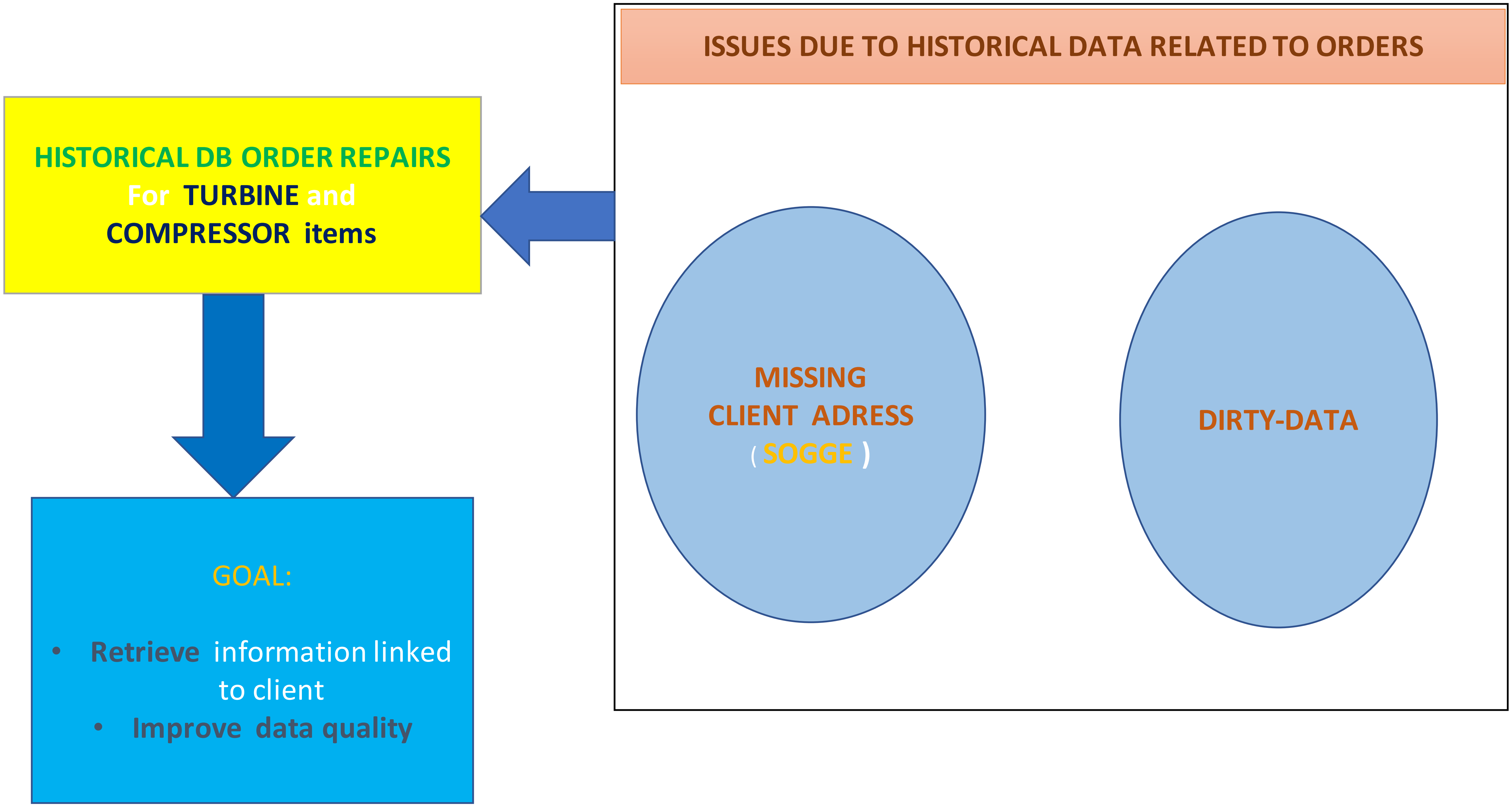
BUSINESS CONTEXT

TURBOMACHINERY& PROCESS SOLUTION

REPAIRS

SPARE PARTS

HISTORICAL DB ORDER REPAIRS :



ISSUES DUE TO HISTORICAL DATA RELATED TO ORDERS

MISSING
CLIENT ADRESS
(SOGGE)

DIRTY-DATA

HISTORICAL DB ORDER REPAIRS
For TURBINE and
COMPRESSOR items

GOAL:

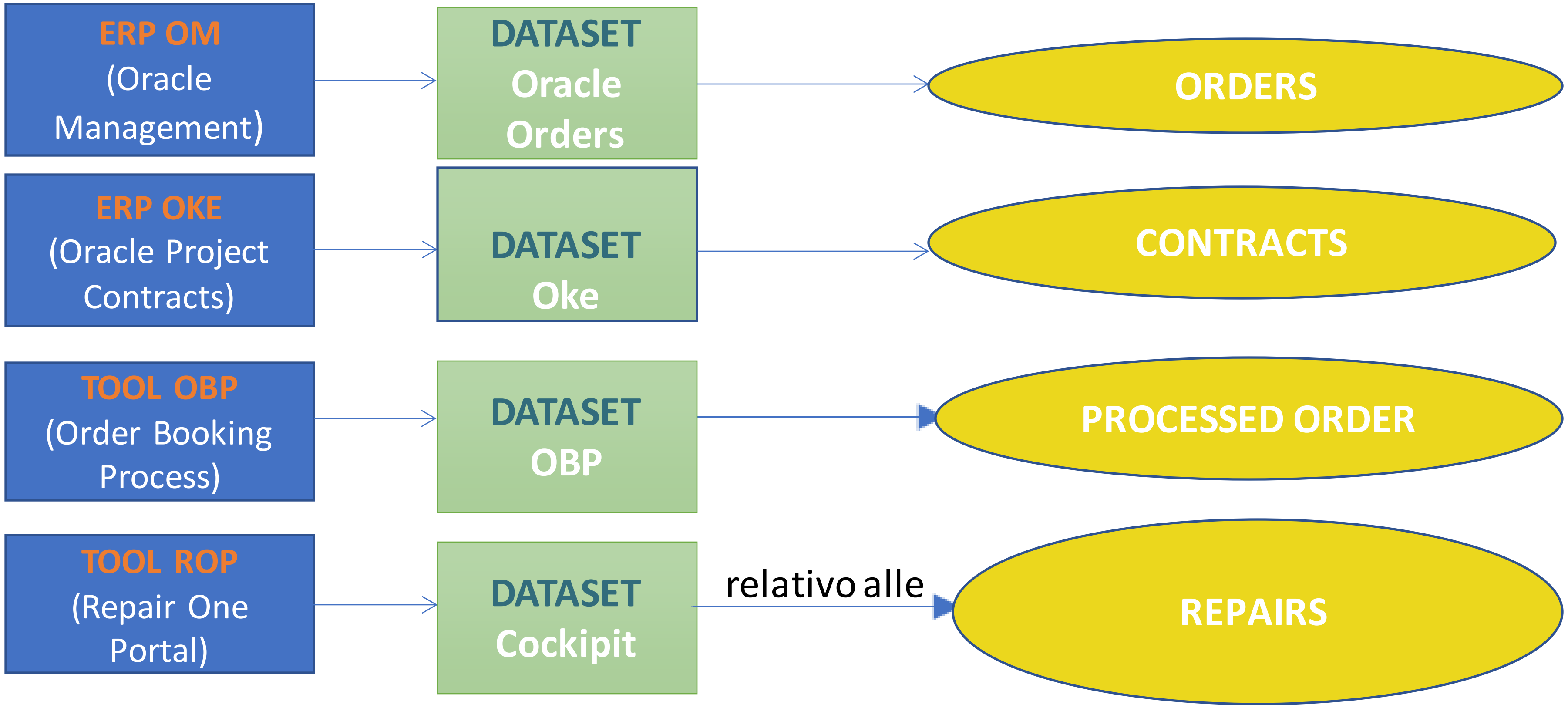
- Retrieve information linked to client
- Improve data quality

HISTORICAL DB ORDER REPAIRS

Combination from multiple DATA SOURCES :

-ERP

-TOOLS



ERP OM (Oracle Management)

DATASET Oracle Orders

ORDERS

ERP OKE (Oracle Project Contracts)

DATASET Oke

CONTRACTS

TOOL OBP (Order Booking Process)

DATASET OBP

PROCESSED ORDER

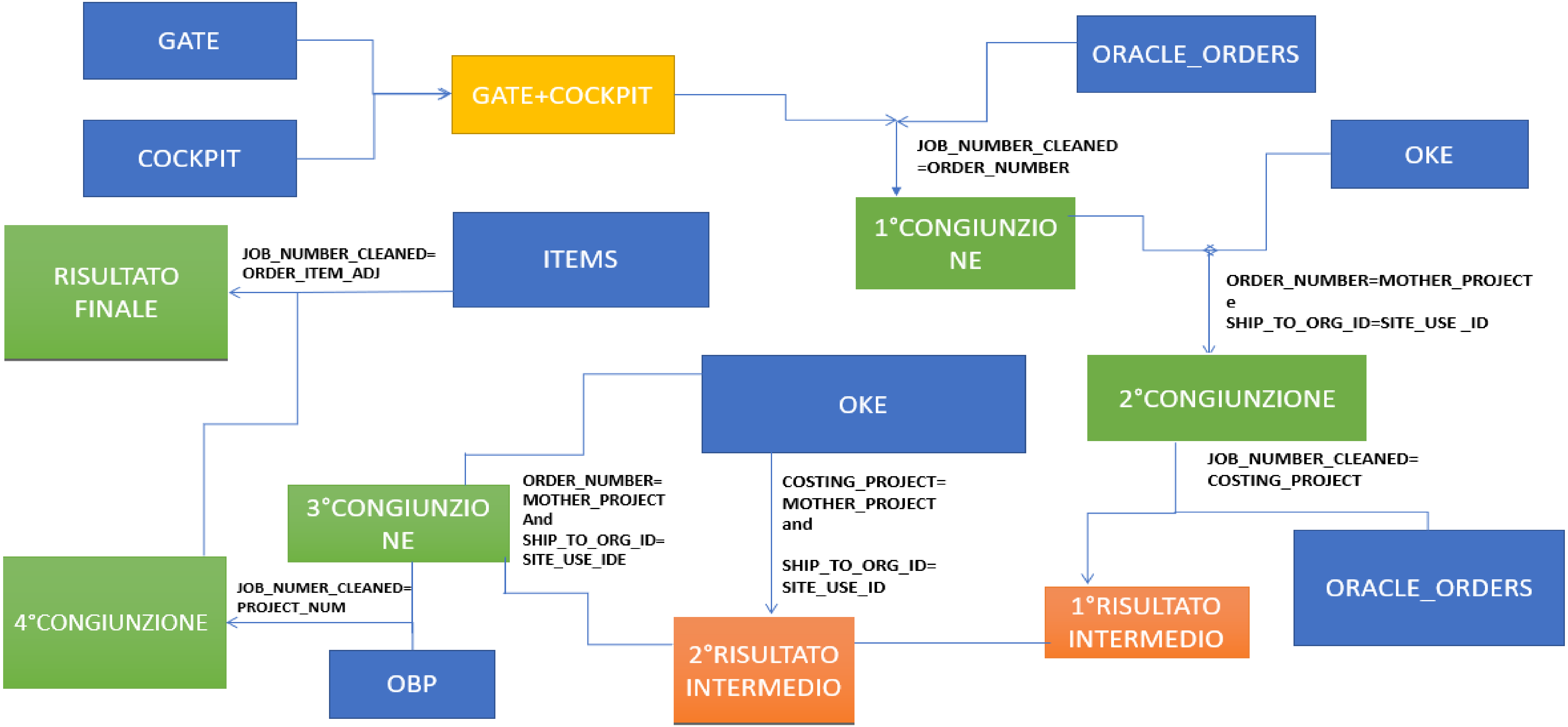
TOOL ROP (Repair One Portal)

DATASET Cockpit

relativo alle

REPAIRS

HISTORICAL DB ORDER REPAIRS


GATE
COCKPIT
GATE+COCKPIT
ORACLE_ORDERS
JOB_NUMBER_CLEANED =ORDER_NUMBER
1°CONGIUNZIONE
OKE
ORDER_NUMBER=MOTHER_PROJECT e SHIP_TO_ORG_ID=SITE_USE _ID
2°CONGIUNZIONE
RISULTATO FINALE
JOB_NUMBER_CLEANED= ORDER_ITEM_ADJ
ITEMS
OKE
JOB_NUMBER_CLEANED= COSTING_PROJECT
ORACLE_ORDERS
1°RISULTATO INTERMEDIO
COSTING_PROJECT= MOTHER_PROJECT and SHIP_TO_ORG_ID= SITE_USE_ID
2°RISULTATO INTERMEDIO
3°CONGIUNZIONE
ORDER_NUMBER= MOTHER_PROJECT And SHIP_TO_ORG_ID= SITE_USE_IDE
4°CONGIUNZIONE
JOB_NUMER_CLEANED= PROJECT_NUM
OBP

HISTORICAL DB ORDER REPAIRS

Main steps:

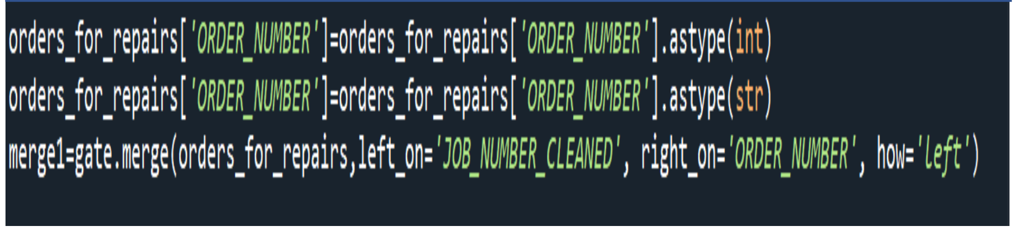
merge(how='left')

- Convert data types keys

- Handle with duplicates and null values
- Data cleaning
- Remove duplicate columns

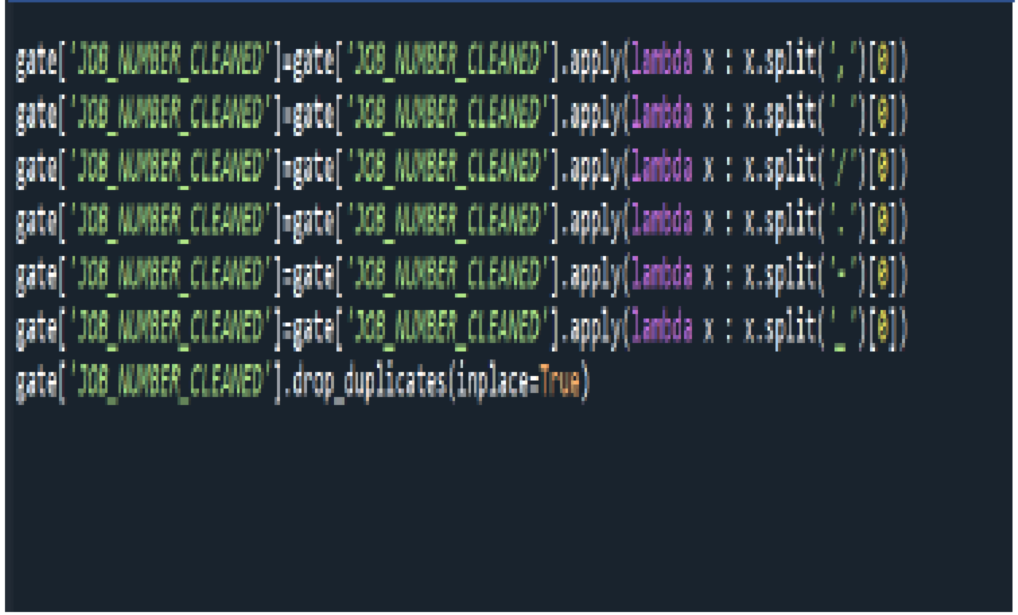
HISTORICAL DB ORDER REPAIRS

COVERTING INTO STRING A DATA TYPE KEY TO ENABLE A COMBINATION



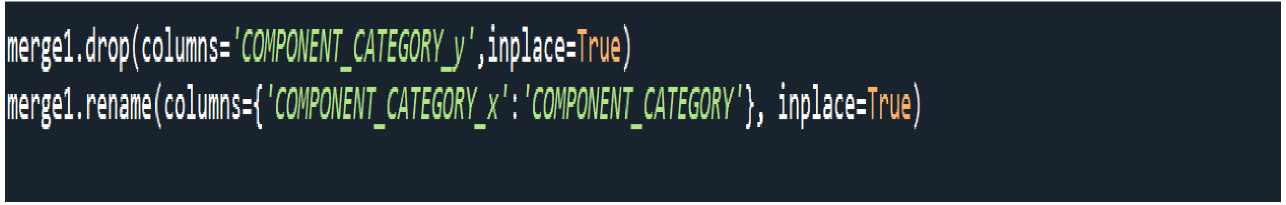
```
orders_for_repairs['ORDER_NUMBER']=orders_for_repairs['ORDER_NUMBER'].astype(int)
orders_for_repairs['ORDER_NUMBER']=orders_for_repairs['ORDER_NUMBER'].astype(str)
merge1=gate.merge(orders_for_repairs,left_on='JOB_NUMBER_CLEANED', right_on='ORDER_NUMBER', how='left')
```

CLEANING DATA AND REMOVING DUPLICATES VALUES



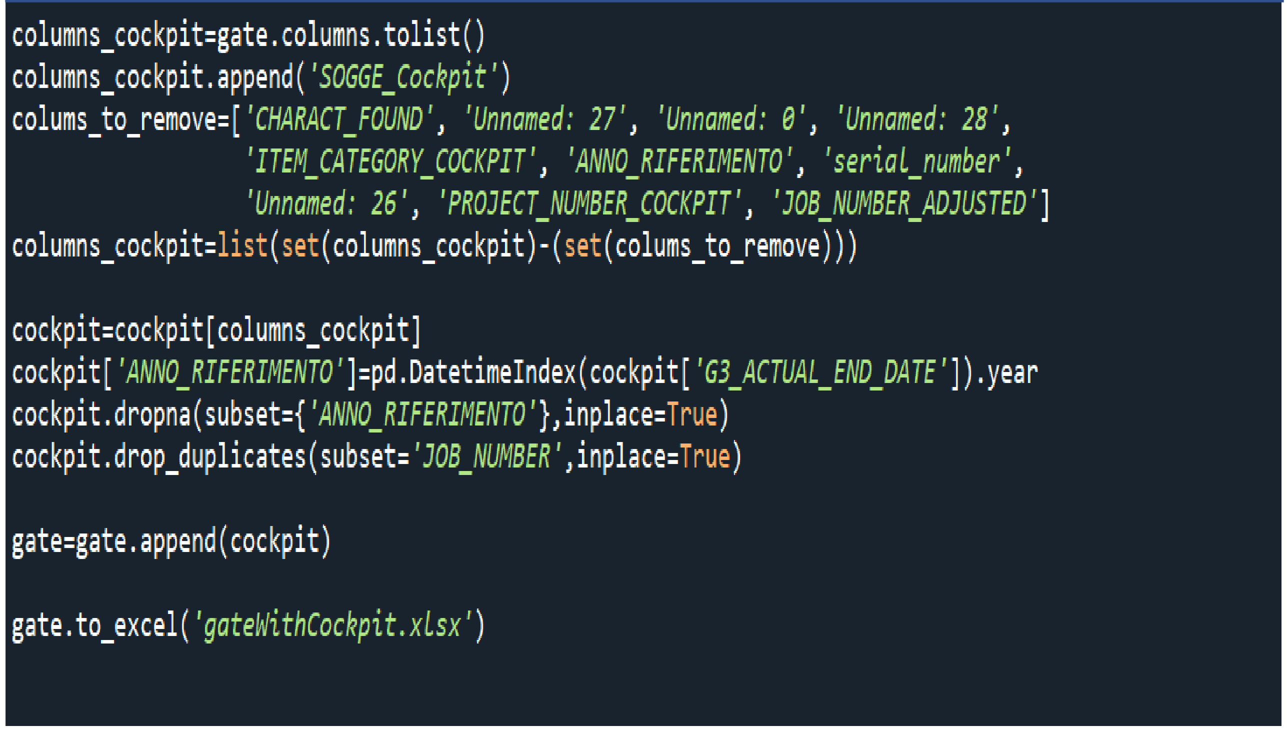
```
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split(',')[0])
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split(' ')[0])
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split('/')[0])
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split('.')[0])
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split('-')[0])
gate['JOB_NUMBER_CLEANED']=gate['JOB_NUMBER_CLEANED'].apply(lambda x : x.split('_')[0])
gate['JOB_NUMBER_CLEANED'].drop_duplicates(inplace=True)
```

REMOVING DUPLICATE COLUMNS



```
merge1.drop(columns='COMPONENT_CATEGORY_y',inplace=True)
merge1.rename(columns={'COMPONENT_CATEGORY_x':'COMPONENT_CATEGORY'}, inplace=True)
```

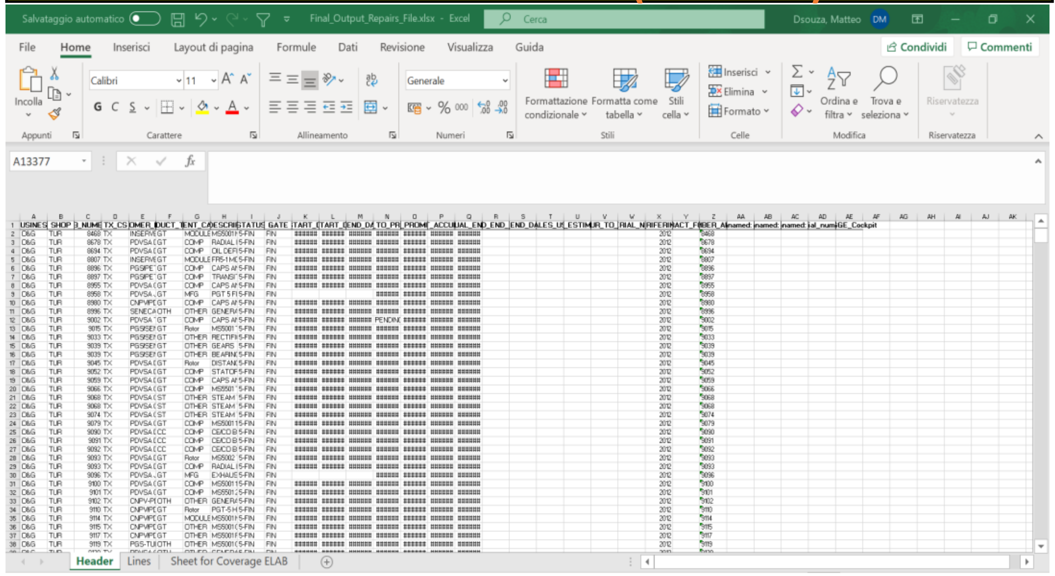
REMOVING ROWS WITH KEY'S NULL VALUES



```
columns_cockpit=gate.columns.tolist()
columns_cockpit.append('SOGGE_Cockpit')
colums_to_remove=['CHARACT_FOUND', 'Unnamed: 27', 'Unnamed: 0', 'Unnamed: 28',
                  'ITEM_CATEGORY_COCKPIT', 'ANNO_RIFERIMENTO', 'serial_number',
                  'Unnamed: 26', 'PROJECT_NUMBER_COCKPIT', 'JOB_NUMBER_ADJUSTED']
columns_cockpit=list(set(columns_cockpit)-(set(colums_to_remove)))

cockpit=cockpit[columns_cockpit]
cockpit['ANNO_RIFERIMENTO']=pd.DatetimeIndex(cockpit['G3_ACTUAL_END_DATE']).year
cockpit.dropna(subset={'ANNO_RIFERIMENTO'},inplace=True)
cockpit.drop_duplicates(subset='JOB_NUMBER',inplace=True)

gate=gate.append(cockpit)

gate.to_excel('gateWithCockpit.xlsx')
```

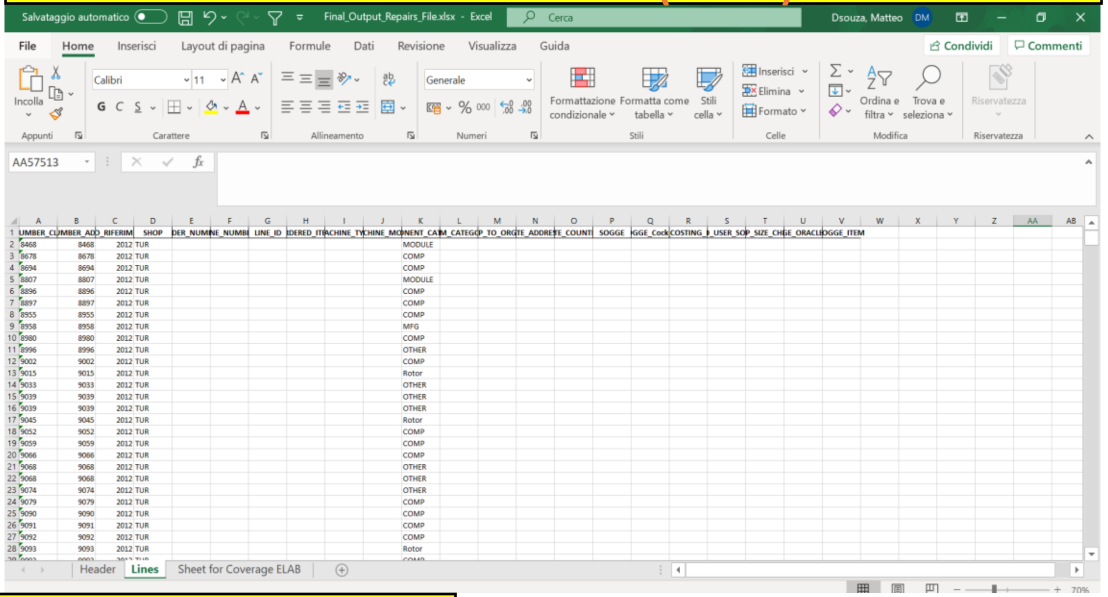
HISTORICAL DB ORDER REPAIRS

LIVELLO DI TESTATA (HEADER)



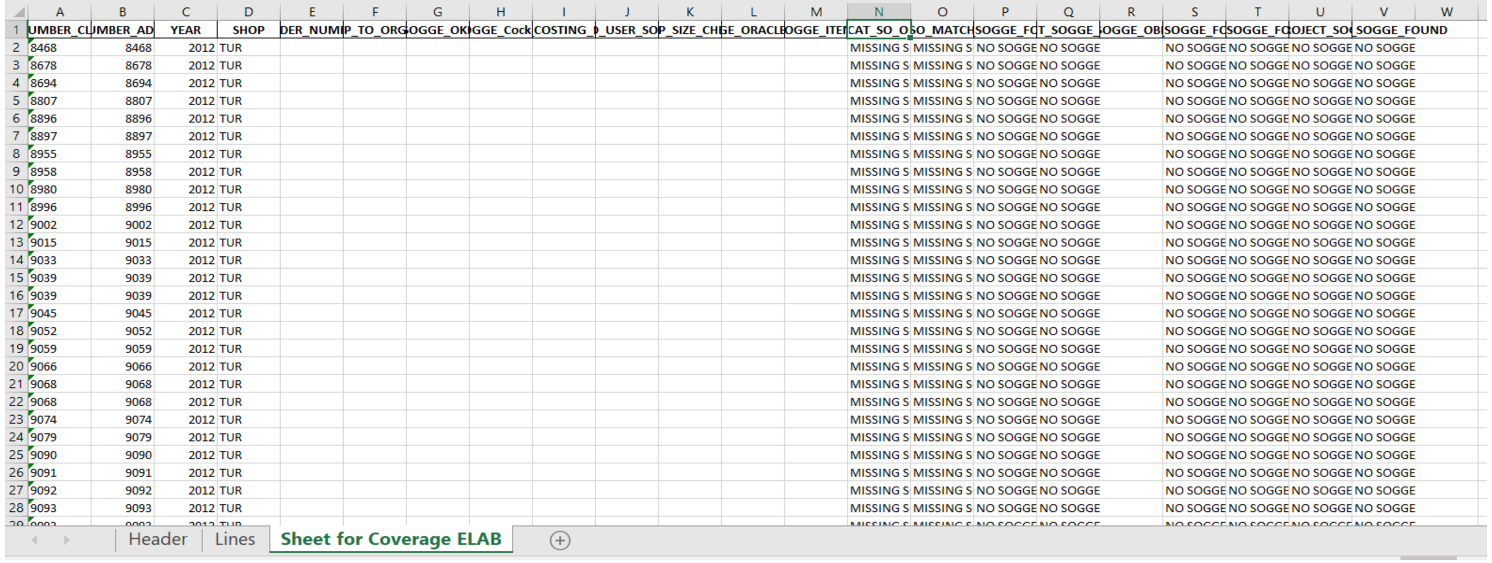
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	USINES	SHOP	B_NUME	TX_CS	OMER_	DUCT_	ENT_CA	DESCRI	STATUS	GATE	TART_	TART_	END_DA	TO_PR	_PROM	_ACCU	UAL_EN	D_END_	END_DA	LES_U	_ESTIM	R_TO_	RIAL_N	RIFERIM	ACT_F	BER_A	nnamed:	nnamed:	nnamed:	ial_num	GE_Cockpit
2	O&G	TUR	8468	TX	INSERV	GT	MODULE	MS5001	5-FIN	FIN	######	######	######	######	######	######	######							2012		8468					
3	O&G	TUR	8678	TX	PDVSA (	GT	COMP	RADIAL I	5-FIN	FIN	######	######	######	######	######	######	######							2012		8678					
4	O&G	TUR	8694	TX	PDVSA (	GT	COMP	OIL DEF	5-FIN	FIN	######	######	######	######	######	######	######							2012		8694					
5	O&G	TUR	8807	TX	INSERV	GT	MODULE	FR5-1 MC	5-FIN	FIN	######	######	######	######	######	######	######							2012		8807					
6	O&G	TUR	8896	TX	PGSIPE	GT	COMP	CAPS A	5-FIN	FIN	######	######	######	######	######	######	######							2012		8896					
7	O&G	TUR	8897	TX	PGSIPE	GT	COMP	TRANSI	5-FIN	FIN	######	######	######	######	######	######	######							2012		8897					
8	O&G	TUR	8955	TX	PDVSA (	GT	COMP	CAPS A	5-FIN	FIN	######	######	######	######	######	######	######							2012		8955					
9	O&G	TUR	8958	TX	PDVSA .	GT	MFG	PGT 5 FI	5-FIN	FIN				######	######	######	######							2012		8958					
10	O&G	TUR	8980	TX	CNPVIPC	GT	COMP	CAPS A	5-FIN	FIN	######	######	######	######	######	######	######							2012		8980					
11	O&G	TUR	8996	TX	SENECA	OTH	OTHER	GENERA	5-FIN	FIN	######	######	######	######	######	######	######							2012		8996					
12	O&G	TUR	9002	TX	PDVSA	GT	COMP	CAPS A	5-FIN	FIN	######	######	######	PENDING	######	######	######							2012		9002					
13	O&G	TUR	9015	TX	PGSISEI	GT	Rotor	MS5001	5-FIN	FIN	######	######	######	######	######	######	######							2012		9015					
14	O&G	TUR	9033	TX	PGSISEI	GT	OTHER	RECTIFI	5-FIN	FIN	######	######	######	######	######	######	######							2012		9033					
15	O&G	TUR	9039	TX	PGSISEI	GT	OTHER	GEARS	5-FIN	FIN	######	######	######	######	######	######	######							2012		9039					
16	O&G	TUR	9039	TX	PGSISEI	GT	OTHER	BEARING	5-FIN	FIN	######	######	######	######	######	######	######							2012		9039					
17	O&G	TUR	9045	TX	PDVSA (	GT	Rotor	DISTAN	5-FIN	FIN	######	######	######	######	######	######	######							2012		9045					
18	O&G	TUR	9052	TX	PDVSA (	GT	COMP	STATOR	5-FIN	FIN	######	######	######	######	######	######	######							2012		9052					
19	O&G	TUR	9059	TX	PDVSA (	GT	COMP	CAPS A	5-FIN	FIN	######	######	######	######	######	######	######							2012		9059					
20	O&G	TUR	9066	TX	PDVSA (	GT	COMP	MS5501	5-FIN	FIN	######	######	######	######	######	######	######							2012		9066					
21	O&G	TUR	9068	TX	PDVSA (	ST	OTHER	STEAM	5-FIN	FIN	######	######	######	######	######	######	######							2012		9068					
22	O&G	TUR	9068	TX	PDVSA (	ST	OTHER	STEAM	5-FIN	FIN	######	######	######	######	######	######	######							2012		9068					
23	O&G	TUR	9074	TX	PDVSA (	ST	OTHER	STEAM	5-FIN	FIN	######	######	######	######	######	######	######							2012		9074					
24	O&G	TUR	9079	TX	PDVSA (	GT	COMP	MS5001 1	5-FIN	FIN	######	######	######	######	######	######	######							2012		9079					
25	O&G	TUR	9090	TX	PDVSA (	CC	COMP	CECO B	5-FIN	FIN	######	######	######	######	######	######	######							2012		9090					
26	O&G	TUR	9091	TX	PDVSA (	CC	COMP	CECO B	5-FIN	FIN	######	######	######	######	######	######	######							2012		9091					
27	O&G	TUR	9092	TX	PDVSA (	CC	COMP	CECO B	5-FIN	FIN	######	######	######	######	######	######	######							2012		9092					
28	O&G	TUR	9093	TX	PDVSA (	GT	Rotor	MS5002	5-FIN	FIN	######	######	######	######	######	######	######							2012		9093					
29	O&G	TUR	9093	TX	PDVSA (	GT	COMP	RADIAL I	5-FIN	FIN	######	######	######	######	######	######	######							2012		9093					
30	O&G	TUR	9096	TX	PDVSA .	GT	MFG	EXHAUS	5-FIN	FIN				######	######	######	######							2012		9096					
31	O&G	TUR	9100	TX	PDVSA (	GT	COMP	MS5001 1	5-FIN	FIN	######	######	######	######	######	######	######							2012		9100					
32	O&G	TUR	9101	TX	PDVSA (	GT	COMP	MS5501 2	5-FIN	FIN	######	######	######	######	######	######	######							2012		9101					
33	O&G	TUR	9102	TX	CNPV-PI	OTH	OTHER	GENERA	5-FIN	FIN	######	######	######	######	######	######	######							2012		9102					
34	O&G	TUR	9110	TX	CNPVIPC	GT	Rotor	PGT-5 H	5-FIN	FIN	######	######	######	######	######	######	######							2012		9110					
35	O&G	TUR	9114	TX	CNPVIPC	GT	MODULE	MS5001 H	5-FIN	FIN	######	######	######	######	######	######	######							2012		9114					
36	O&G	TUR	9115	TX	CNPVIPC	GT	OTHER	MS5001 (	5-FIN	FIN	######	######	######	######	######	######	######							2012		9115					
37	O&G	TUR	9117	TX	CNPVIPC	GT	OTHER	MS5001 F	5-FIN	FIN	######	######	######	######	######	######	######							2012		9117					
38	O&G	TUR	9119	TX	PGS-TUI	OTH	OTHER	MS5001 (	5-FIN	FIN	######	######	######	######	######	######	######							2012		9119					

LIVELLO DI LINEE (LINES)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	UMBER_CL	UMBER_AD	_RIFERIM	SHOP	DER_NUM	NE_NUMB	LINE_ID	DERED_IT	ACHINE_TY	CHINE_MO	NENT_CA	M_CATEG	P_TO_ORG	TE_ADDRE	E_COUNT	SOGGE	GGE_Cock	COSTING_	_USER_SO	P_SIZE_CH	GE_ORACL	OGGE_ITEM
2	8468	8468	2012	TUR							MODULE											
3	8678	8678	2012	TUR							COMP											
4	8694	8694	2012	TUR							COMP											
5	8807	8807	2012	TUR							MODULE											
6	8896	8896	2012	TUR							COMP											
7	8897	8897	2012	TUR							COMP											
8	8955	8955	2012	TUR							COMP											
9	8958	8958	2012	TUR							MFG											
10	8980	8980	2012	TUR							COMP											
11	8996	8996	2012	TUR							OTHER											
12	9002	9002	2012	TUR							COMP											
13	9015	9015	2012	TUR							Rotor											
14	9033	9033	2012	TUR							OTHER											
15	9039	9039	2012	TUR							OTHER											
16	9039	9039	2012	TUR							OTHER											
17	9045	9045	2012	TUR							Rotor											
18	9052	9052	2012	TUR							COMP											
19	9059	9059	2012	TUR							COMP											
20	9066	9066	2012	TUR							COMP											
21	9068	9068	2012	TUR							OTHER											
22	9068	9068	2012	TUR							OTHER											
23	9074	9074	2012	TUR							OTHER											
24	9079	9079	2012	TUR							COMP											
25	9090	9090	2012	TUR							COMP											
26	9091	9091	2012	TUR							COMP											
27	9092	9092	2012	TUR							COMP											
28	9093	9093	2012	TUR							Rotor											

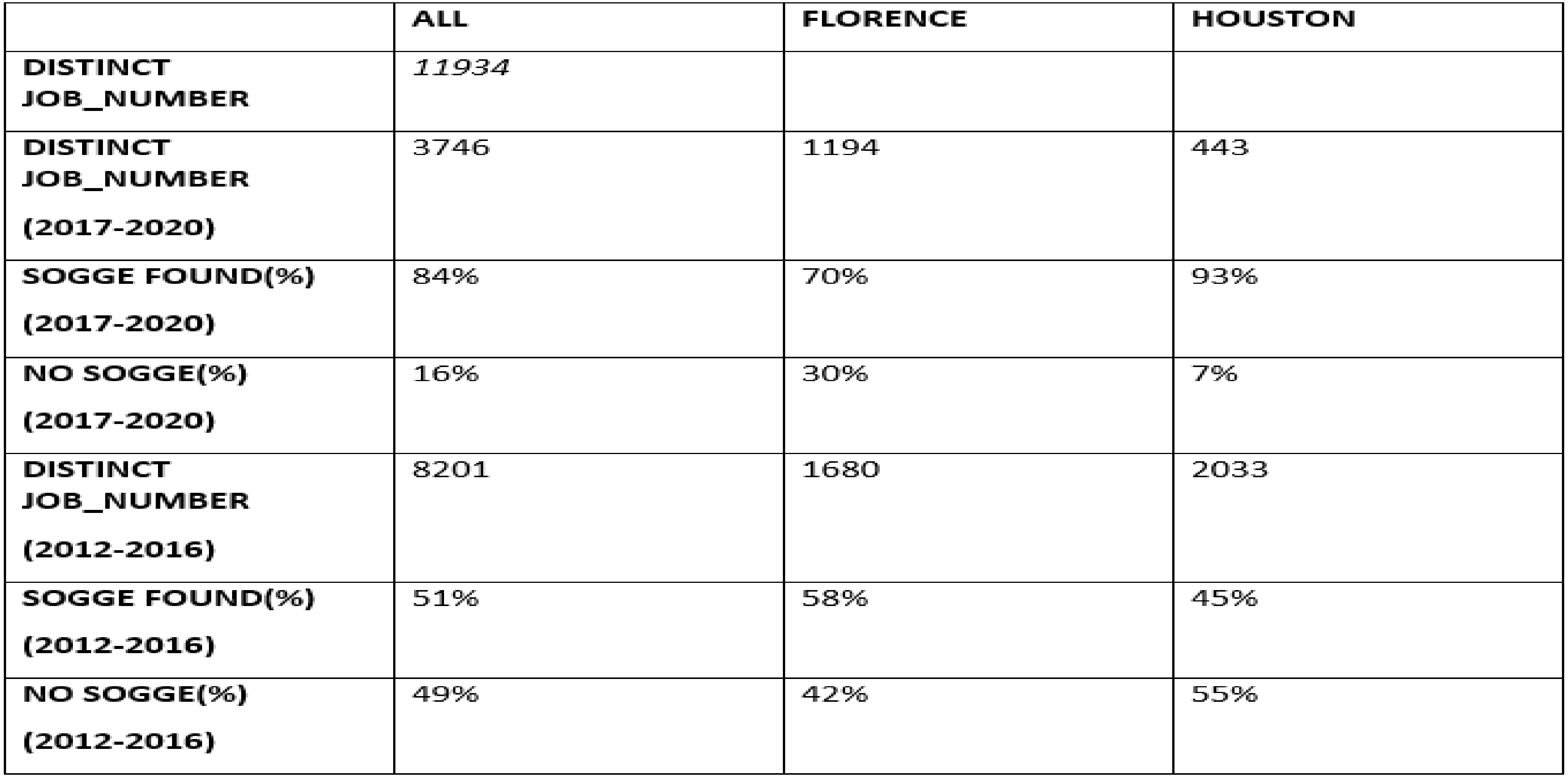
LIVELLO DI COPERTURA (SHEET_FOR_COVERAGE_ELAB)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	UMBER_CL	UMBER_AD	YEAR	SHOP	DER_NUM	P_TO_ORG	OGGE_OK	GGE_Cock	COSTING_	_USER_SO	P_SIZE_CH	E_ORACL	OGGE_ITE	CAT_SO_O	SO_MATCH	SOGGE_FO	T_SOGGE_	OGGE_OB	SOGGE_F	SOGGE_F	OJECT_SO	SOGGE_FOUND
2	8468	8468	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
3	8678	8678	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
4	8694	8694	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
5	8807	8807	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
6	8896	8896	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
7	8897	8897	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
8	8955	8955	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
9	8958	8958	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
10	8980	8980	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
11	8996	8996	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
12	9002	9002	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
13	9015	9015	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
14	9033	9033	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
15	9039	9039	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
16	9039	9039	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
17	9045	9045	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
18	9052	9052	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
19	9059	9059	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
20	9066	9066	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
21	9068	9068	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
22	9068	9068	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
23	9074	9074	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
24	9079	9079	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
25	9090	9090	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
26	9091	9091	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
27	9092	9092	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE
28	9093	9093	2012	TUR										MISSING S	MISSING S	NO SOGGE	NO SOGGE		NO SOGGE	NO SOGGE	NO SOGGE	NO SOGGE

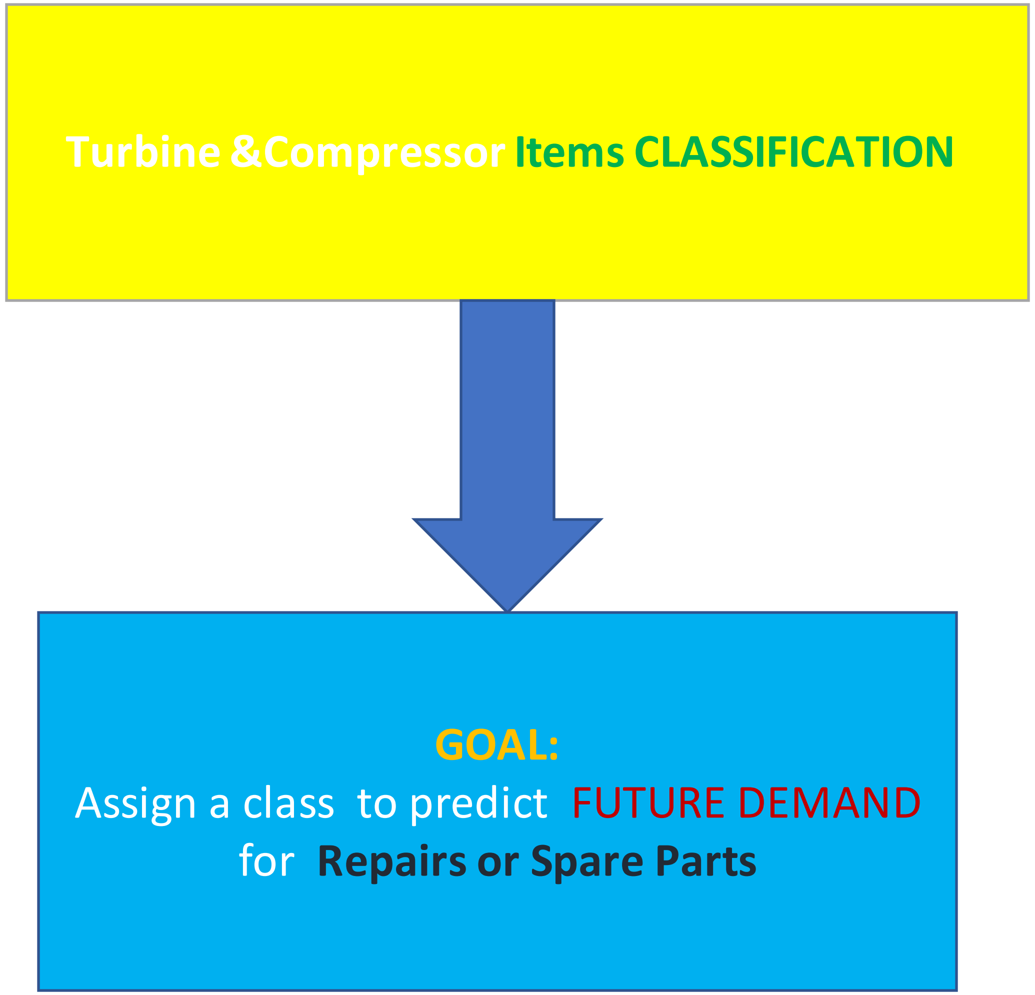
HISTORICAL DB ORDER REPAIRS

- Significant increasing in COVERAGE of client address (SOGGE)



	ALL	FLORENCE	HOUSTON
DISTINCT JOB_NUMBER	*11934*		
DISTINCT JOB_NUMBER (2017-2020)	3746	1194	443
SOGGE FOUND(%) (2017-2020)	84%	70%	93%
NO SOGGE(%) (2017-2020)	16%	30%	7%
DISTINCT JOB_NUMBER (2012-2016)	8201	1680	2033
SOGGE FOUND(%) (2012-2016)	51%	58%	45%
NO SOGGE(%) (2012-2016)	49%	42%	55%

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS



Turbine &Compressor Items CLASSIFICATION

GOAL:
Assign a class to predict FUTURE DEMAND
for **Repairs or Spare Parts**

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS

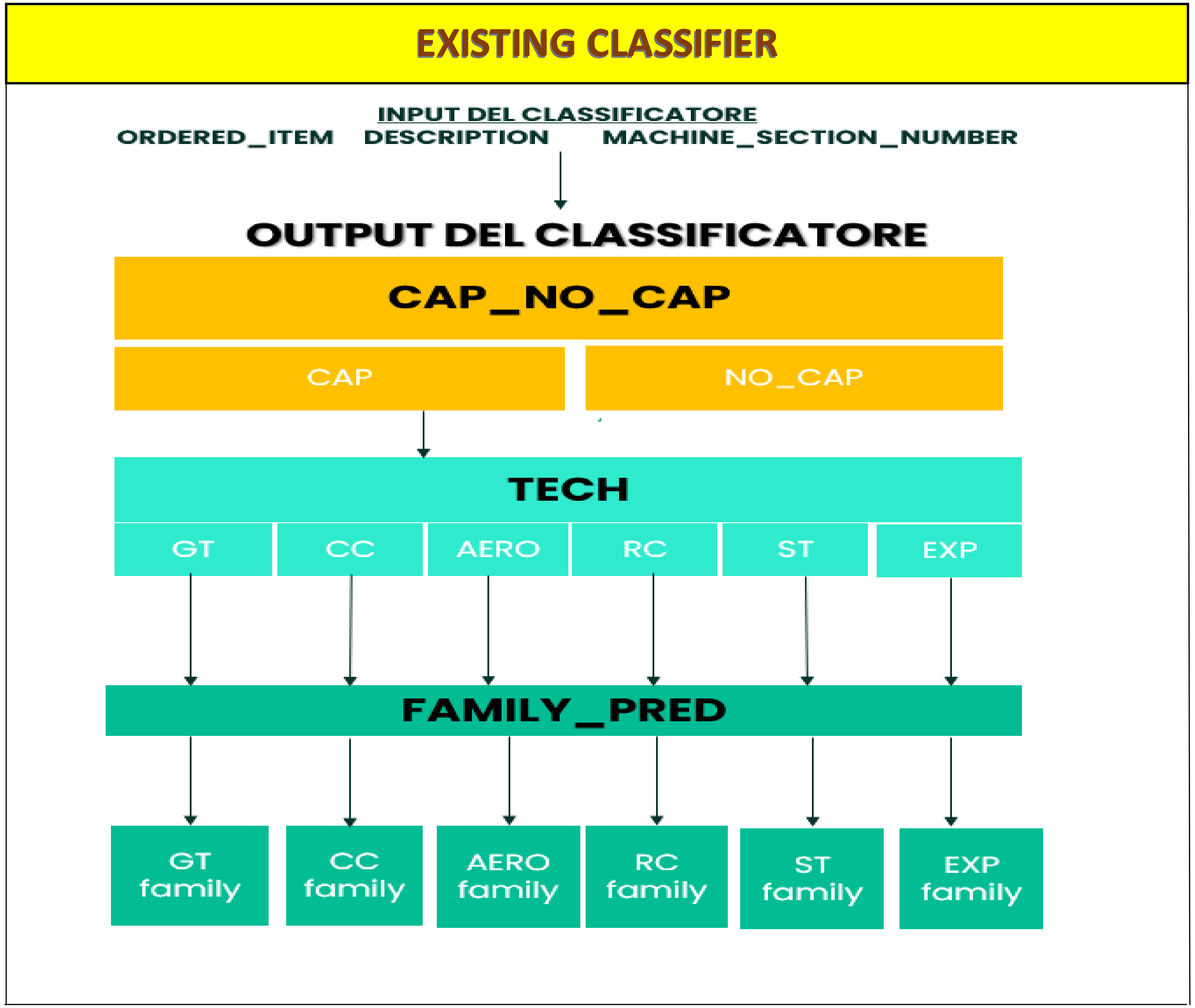
MAIN ISSUE

Uncorrect predictions for some types of classifier's input_

KEY SOLUTION_

ETL Data pipeline:

pre-processing data :



EXISTING CLASSIFIER

INPUT DEL CLASSIFICATORE

ORDERED_ITEM DESCRIPTION MACHINE_SECTION_NUMBER

OUTPUT DEL CLASSIFICATORE

CAP_NO_CAP

CAP

NO_CAP

TECH

GT CC AERO RC ST EXP

FAMILY_PRED

GT family

CC family

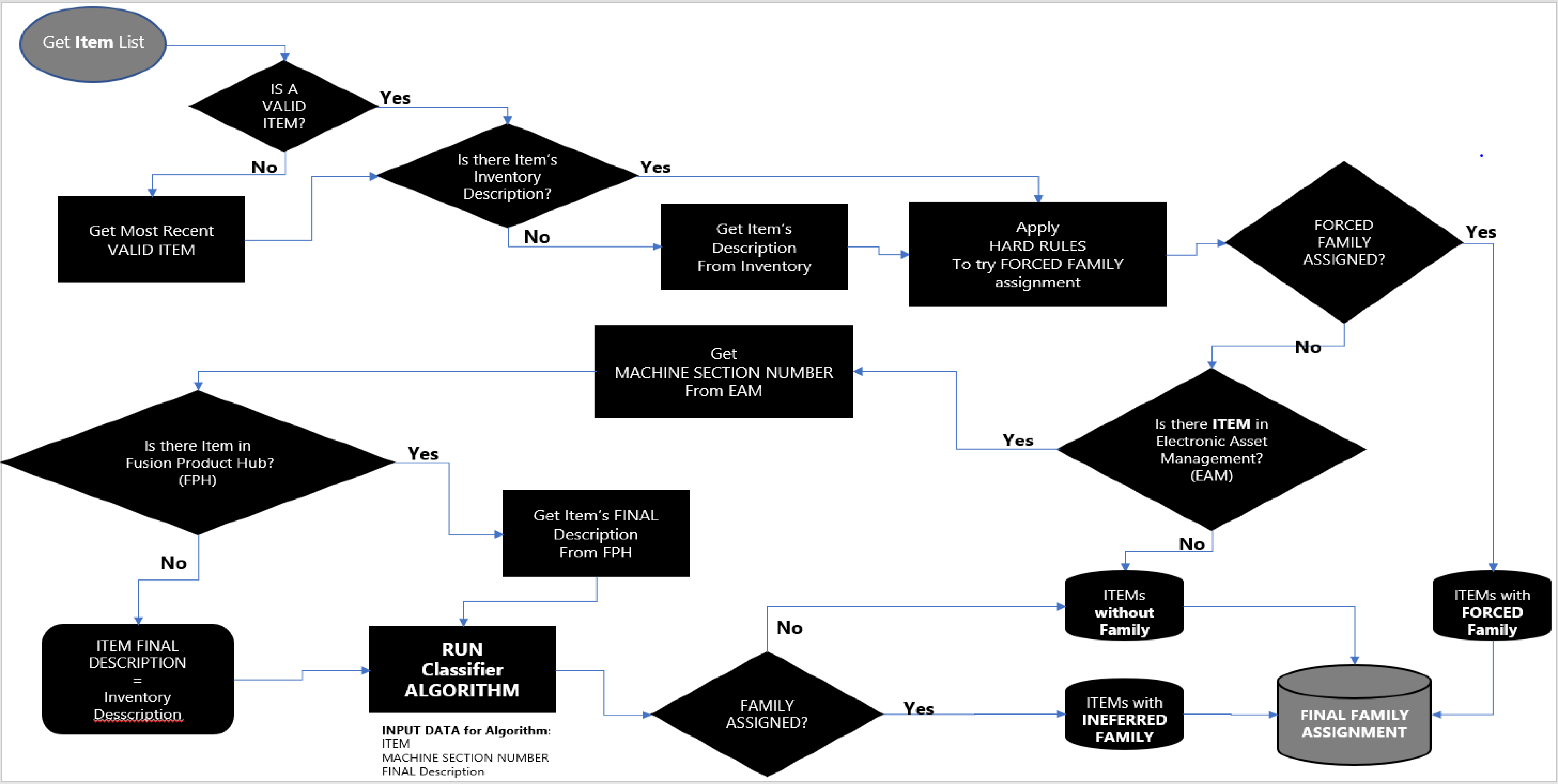
AERO family

RC family

ST family

EXP family

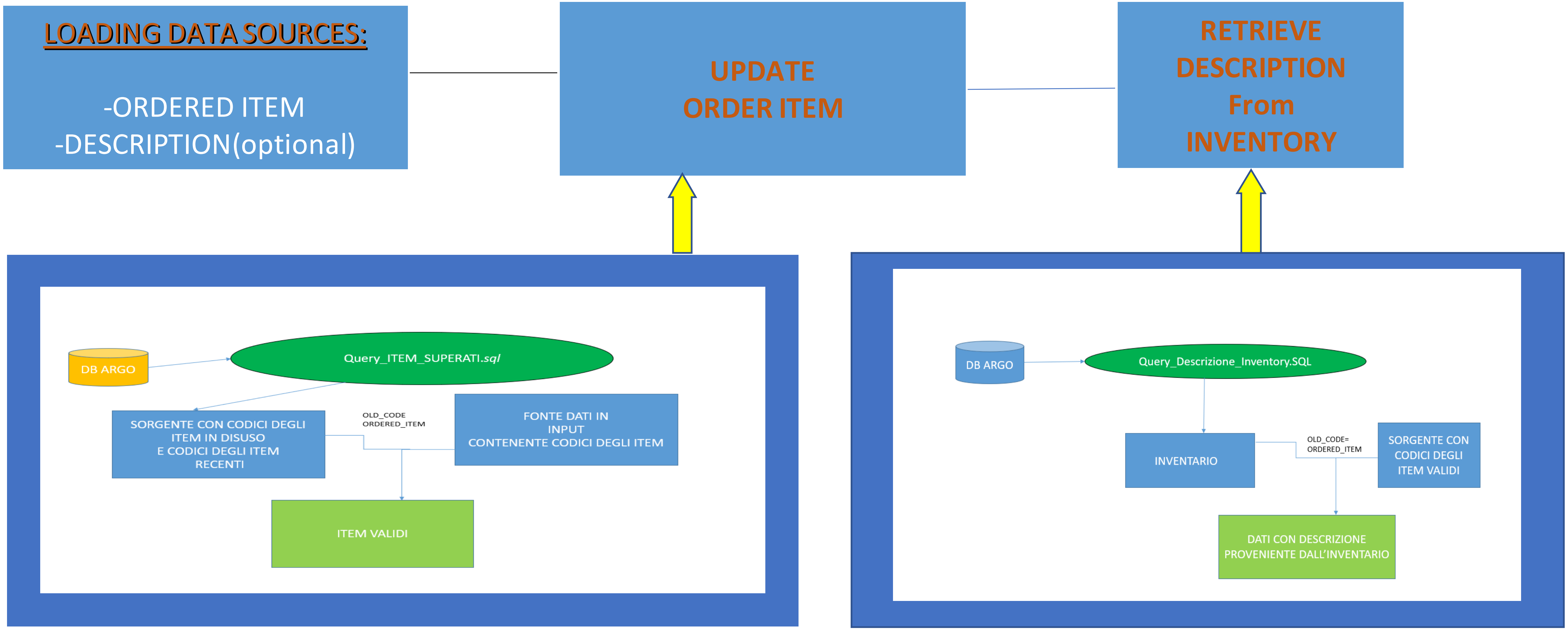
CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS



Get Item List
IS A VALID ITEM?
Yes
No
Get Most Recent VALID ITEM
Is there Item's Inventory Description?
Yes
No
Get Item's Description From Inventory
Apply HARD RULES To try FORCED FAMILY assignment
FORCED FAMILY ASSIGNED?
Yes
No
Is there ITEM in Electronic Asset Management? (EAM)
Yes
No
Get MACHINE SECTION NUMBER From EAM
Is there Item in Fusion Product Hub? (FPH)
Yes
No
Get Item's FINAL Description From FPH
ITEM FINAL DESCRIPTION = Inventory Desscription
RUN Classifier ALGORITHM
FAMILY ASSIGNED?
No
Yes
ITEMs without Family
ITEMs with INEFERRED FAMILY
ITEMs with FORCED Family
FINAL FAMILY ASSIGNMENT

INPUT DATA for Algorithm:
ITEM
MACHINE SECTION NUMBER
FINAL Description

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS



LOADING DATA SOURCES:

-ORDERED ITEM
-DESCRIPTION(optional)

UPDATE
ORDER ITEM

RETRIEVE
DESCRIPTION
From
INVENTORY

DB ARGO

Query_ITEM_SUPERATI.*sql*

SORGENTE CON CODICI DEGLI ITEM IN DISUSO E CODICI DEGLI ITEM RECENTI

OLD_CODE
ORDERED_ITEM

FONTE DATI IN INPUT CONTENENTE CODICI DEGLI ITEM

ITEM VALIDI

DB ARGO

Query_Descrizione_Inventory.SQL

INVENTARIO

OLD_CODE=
ORDERED_ITEM

SORGENTE CON CODICI DEGLI ITEM VALIDI

DATI CON DESCRIZIONE PROVENIENTE DALL'INVENTARIO

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS

BUSINESS LOGICS :

-identify (**NO_CAP**) items

-assign a FAMILY

DATI RELATIVI A COMPONENTI CHE VENGONO CLASSIFICATI SECONDO LE REGOLE DETERMINISTICHE

COMPONENTI APPARTENENTI AD UNA FAMIGLIA AUSILIARIA

PREFIXES OF ORDERED_ITEM
[I, V]

FAMILY_PREDICTION=
AUX

FAMILY_PRED_EXPLANATION
=
HARD_RULES

CAP_NO_CAP=
NO_CAP

COMPONENTI APPARTENENTI AD UNA FAMIGLIA ALTERNATIVA RISPETTO A QUELLE PREVISTE DAL CLASSIFICATORE

PREFIXES OF ORDERED_ITEM
[N,X,Y,1,C]

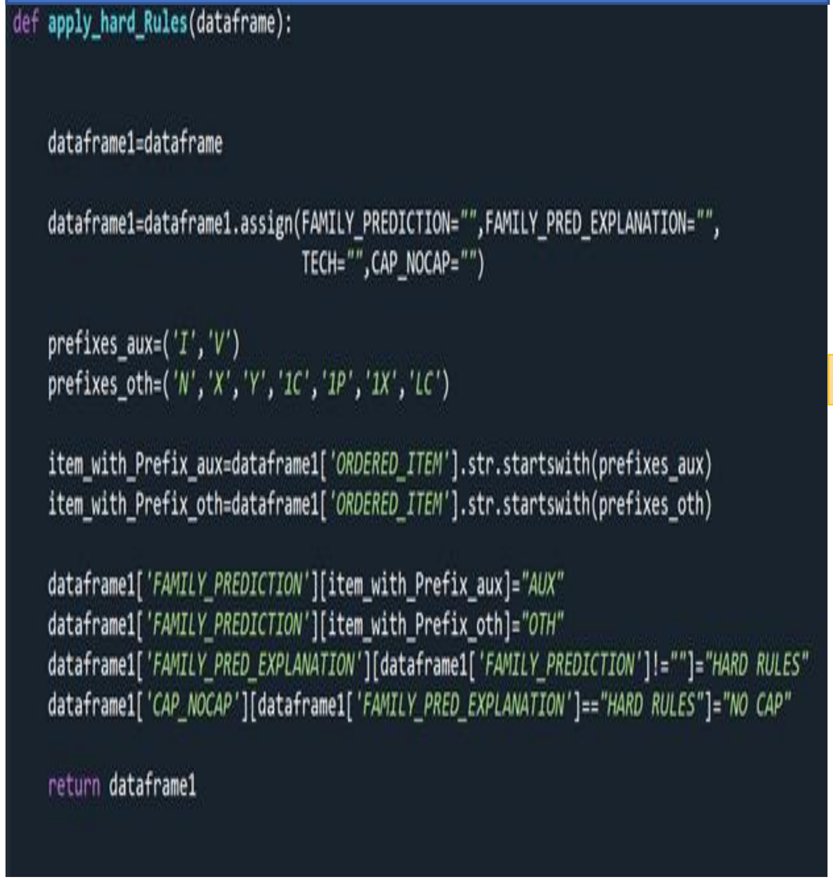
FAMILY_PREDICTION=
OTH

FAMILY_PRED_EXPLANATION
=
HARD_RULES

CAP_NO_CAP=
NO_CAP

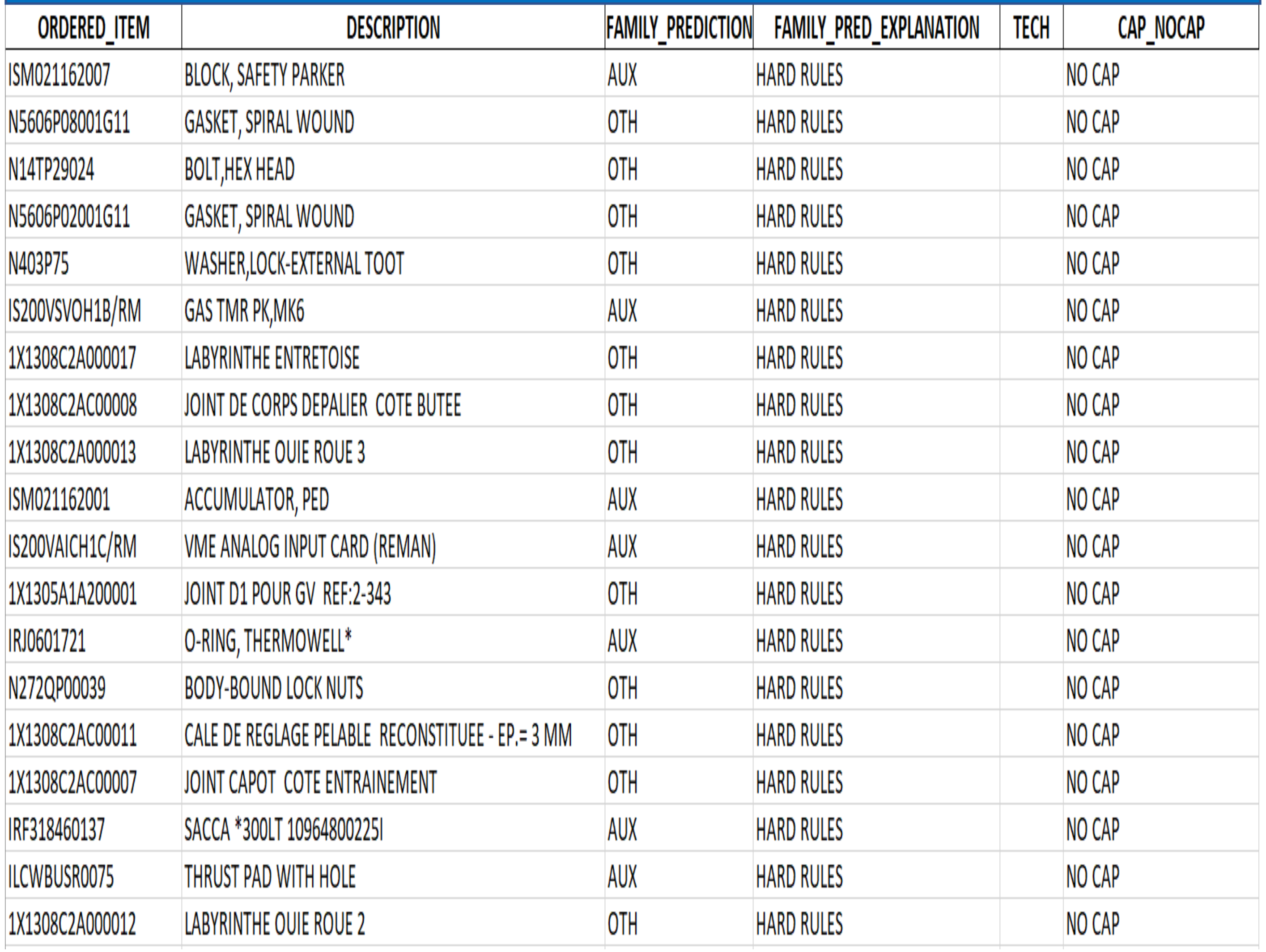
CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS

PYTHON IMPLEMENTATION OF BUSINESS HARD RULES



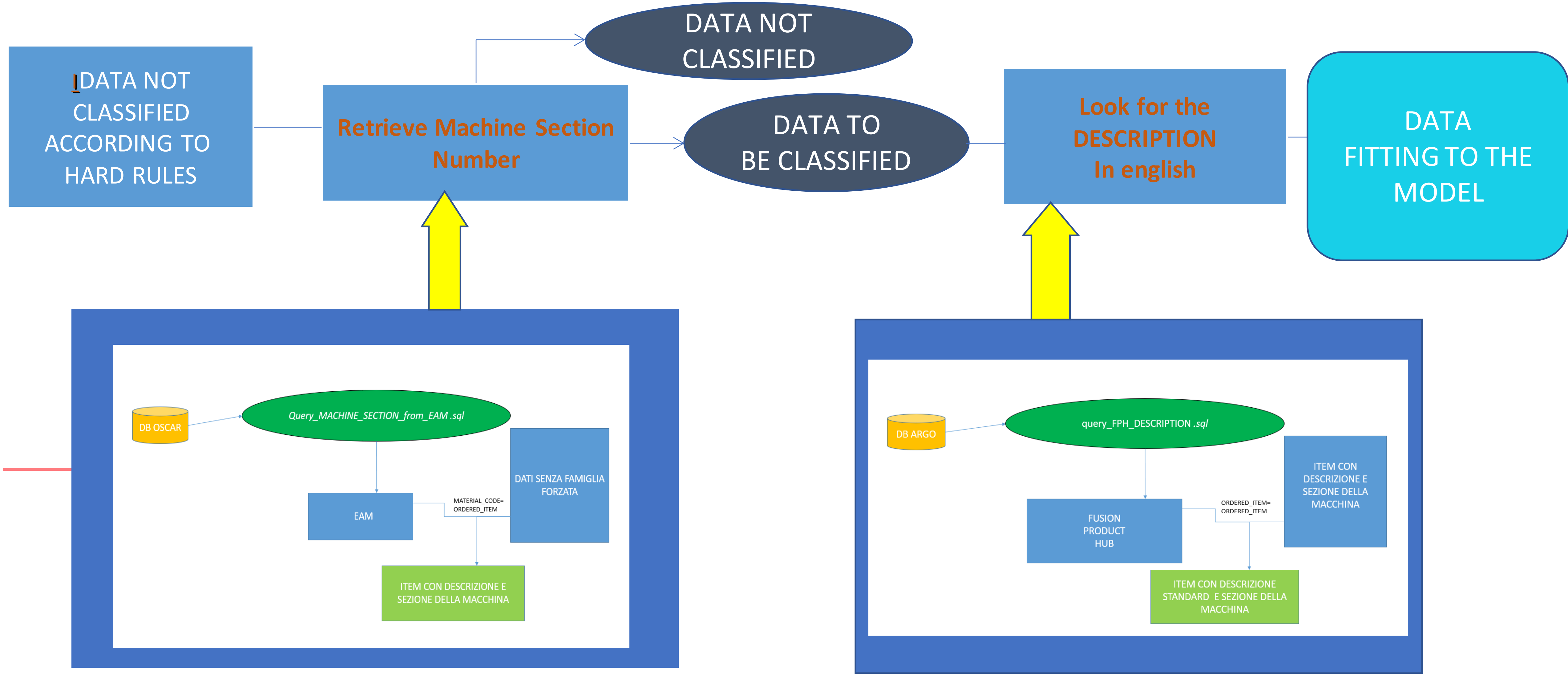
```
def apply_hard_Rules(dataframe):


    dataframe1=dataframe

    dataframe1=dataframe1.assign(FAMILY_PREDICTION="",FAMILY_PRED_EXPLANATION="",
                                 TECH="",CAP_NOCAP="")

    prefixes_aux=('I','V')
    prefixes_oth=('N','X','Y','1C','1P','1X','LC')

    item_with_Prefix_aux=dataframe1['ORDERED_ITEM'].str.startswith(prefixes_aux)
    item_with_Prefix_oth=dataframe1['ORDERED_ITEM'].str.startswith(prefixes_oth)

    dataframe1['FAMILY_PREDICTION'][item_with_Prefix_aux]="AUX"
    dataframe1['FAMILY_PREDICTION'][item_with_Prefix_oth]="OTH"
    dataframe1['FAMILY_PRED_EXPLANATION'][dataframe1['FAMILY_PREDICTION']!=""]="HARD RULES"
    dataframe1['CAP_NOCAP'][dataframe1['FAMILY_PRED_EXPLANATION']=="HARD RULES"]="NO CAP"

    return dataframe1
```

EXAMPLE OF CLASSIFICATION MADE BY HARD RULES BASED ON BUSINESS LOGICS



ORDERED_ITEM	DESCRIPTION	FAMILY_PREDICTION	FAMILY_PRED_EXPLANATION	TECH	CAP_NOCAP
ISM021162007	BLOCK, SAFETY PARKER	AUX	HARD RULES		NO CAP
N5606P08001G11	GASKET, SPIRAL WOUND	OTH	HARD RULES		NO CAP
N14TP29024	BOLT,HEX HEAD	OTH	HARD RULES		NO CAP
N5606P02001G11	GASKET, SPIRAL WOUND	OTH	HARD RULES		NO CAP
N403P75	WASHER,LOCK-EXTERNAL TOOT	OTH	HARD RULES		NO CAP
IS200VSVOH1B/RM	GAS TMR PK,MK6	AUX	HARD RULES		NO CAP
1X1308C2A000017	LABYRINTHE ENTRETOISE	OTH	HARD RULES		NO CAP
1X1308C2AC00008	JOINT DE CORPS DEPALIER COTE BUTEE	OTH	HARD RULES		NO CAP
1X1308C2A000013	LABYRINTHE OUIE ROUE 3	OTH	HARD RULES		NO CAP
ISM021162001	ACCUMULATOR, PED	AUX	HARD RULES		NO CAP
IS200VAICH1C/RM	VME ANALOG INPUT CARD (REMAN)	AUX	HARD RULES		NO CAP
1X1305A1A200001	JOINT D1 POUR GV REF:2-343	OTH	HARD RULES		NO CAP
IRJ0601721	O-RING, THERMOWELL*	AUX	HARD RULES		NO CAP
N272QP00039	BODY-BOUND LOCK NUTS	OTH	HARD RULES		NO CAP
1X1308C2AC00011	CALE DE REGLAGE PELABLE RECONSTITUEE - EP.= 3 MM	OTH	HARD RULES		NO CAP
1X1308C2AC00007	JOINT CAPOT COTE ENTRAINEMENT	OTH	HARD RULES		NO CAP
IRF318460137	SACCA *300LT 10964800225I	AUX	HARD RULES		NO CAP
ILCWBUSR0075	THRUST PAD WITH HOLE	AUX	HARD RULES		NO CAP
1X1308C2A000012	LABYRINTHE OUIE ROUE 2	OTH	HARD RULES		NO CAP

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS



DATA NOT CLASSIFIED ACCORDING TO HARD RULES

Retrieve Machine Section Number

DATA NOT CLASSIFIED

DATA TO BE CLASSIFIED

Look for the DESCRIPTION In english

DATA FITTING TO THE MODEL

DB OSCAR

Query_MACHINE_SECTION_from_EAM .sql

EAM

MATERIAL_CODE= ORDERED_ITEM

DATI SENZA FAMIGLIA FORZATA

ITEM CON DESCRIZIONE E SEZIONE DELLA MACCHINA

DB ARGO

query_FPH_DESCRIPTION *.sql*

FUSION PRODUCT HUB

ORDERED_ITEM= ORDERED_ITEM

ITEM CON DESCRIZIONE E SEZIONE DELLA MACCHINA

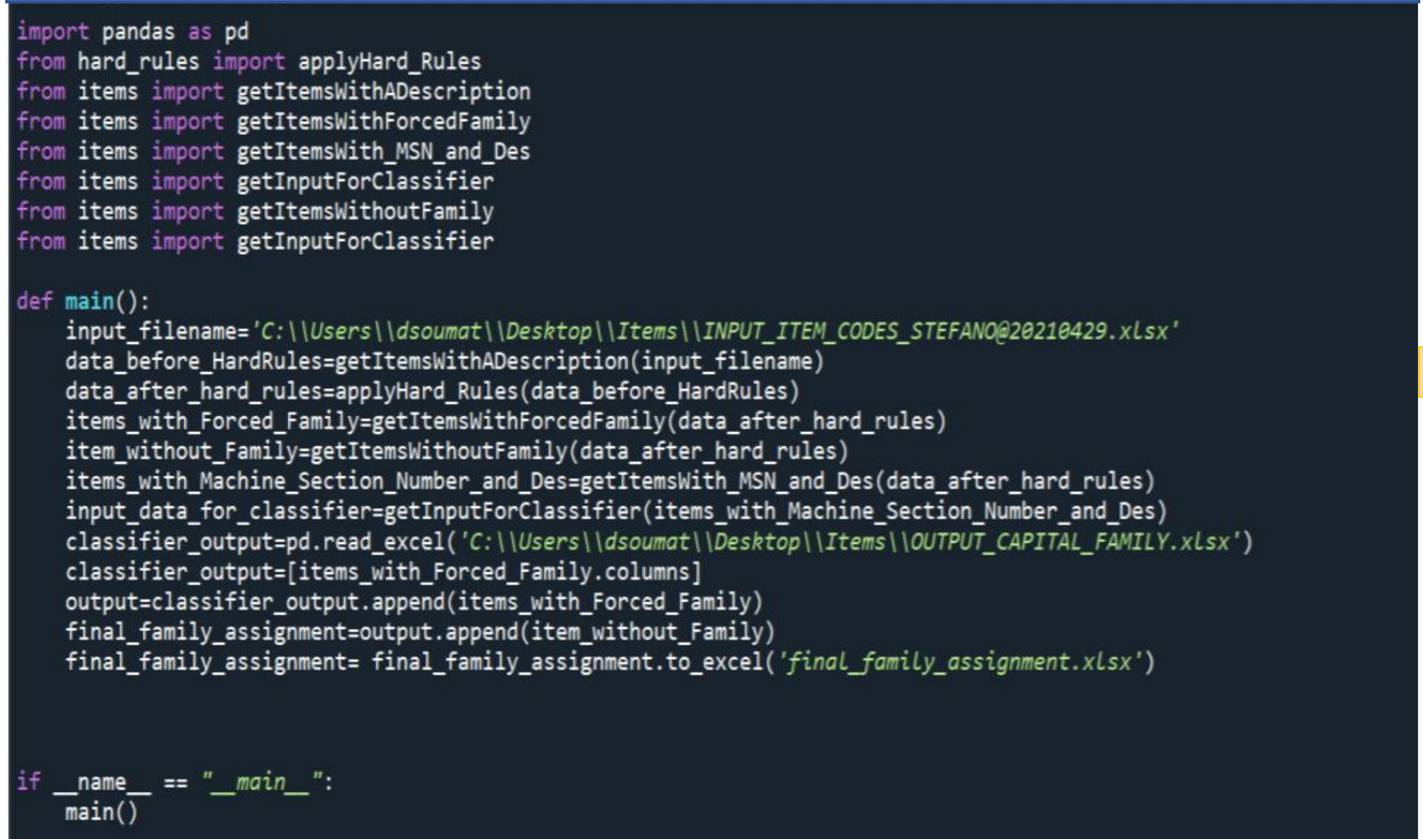
ITEM CON DESCRIZIONE STANDARD E SEZIONE DELLA MACCHINA

CLASSIFICATION OF TURBINE AND COMPRESSOR ITEMS

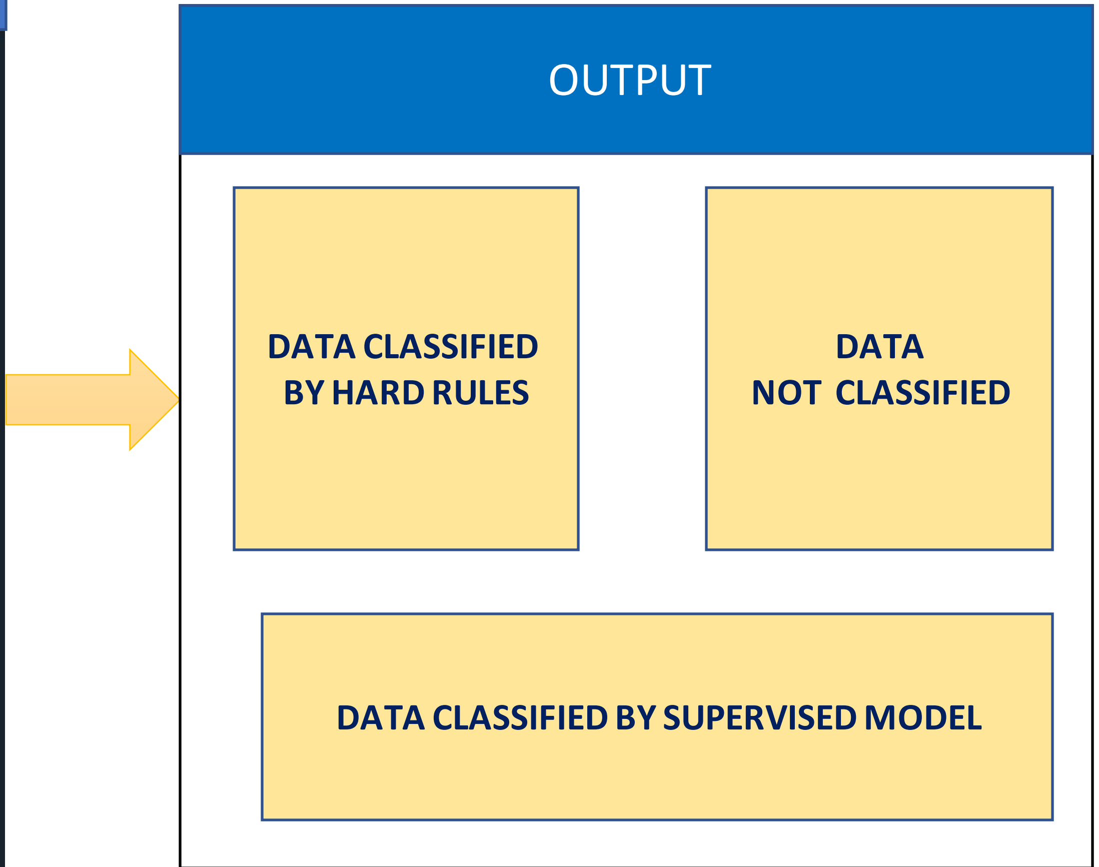
-SPEED UP the classification task

-IMPROVE **the outcome**

DATA ETL PIPELINE OVERFLOW IMPLEMENTATION



```
import pandas as pd
from hard_rules import applyHard_Rules
from items import getItemsWithADescription
from items import getItemsWithForcedFamily
from items import getItemsWith_MSN_and_Des
from items import getInputForClassifier
from items import getItemsWithoutFamily
from items import getInputForClassifier

def main():
    input_filename='C:\\Users\\dsoumat\\Desktop\\Items\\INPUT_ITEM_CODES_STEFANO@20210429.xlsx'
    data_before_HardRules=getItemsWithADescription(input_filename)
    data_after_hard_rules=applyHard_Rules(data_before_HardRules)
    items_with_Forced_Family=getItemsWithForcedFamily(data_after_hard_rules)
    item_without_Family=getItemsWithoutFamily(data_after_hard_rules)
    items_with_Machine_Section_Number_and_Des=getItemsWith_MSN_and_Des(data_after_hard_rules)
    input_data_for_classifier=getInputForClassifier(items_with_Machine_Section_Number_and_Des)
    classifier_output=pd.read_excel('C:\\Users\\dsoumat\\Desktop\\Items\\OUTPUT_CAPITAL_FAMILY.xlsx')
    classifier_output=[items_with_Forced_Family.columns]
    output=classifier_output.append(items_with_Forced_Family)
    final_family_assignment=output.append(item_without_Family)
    final_family_assignment= final_family_assignment.to_excel('final_family_assignment.xlsx')



if __name__ == "__main__":
    main()
```



OUTPUT

DATA CLASSIFIED BY HARD RULES

DATA NOT CLASSIFIED

DATA CLASSIFIED BY SUPERVISED MODEL