



Let's Talk Data Science

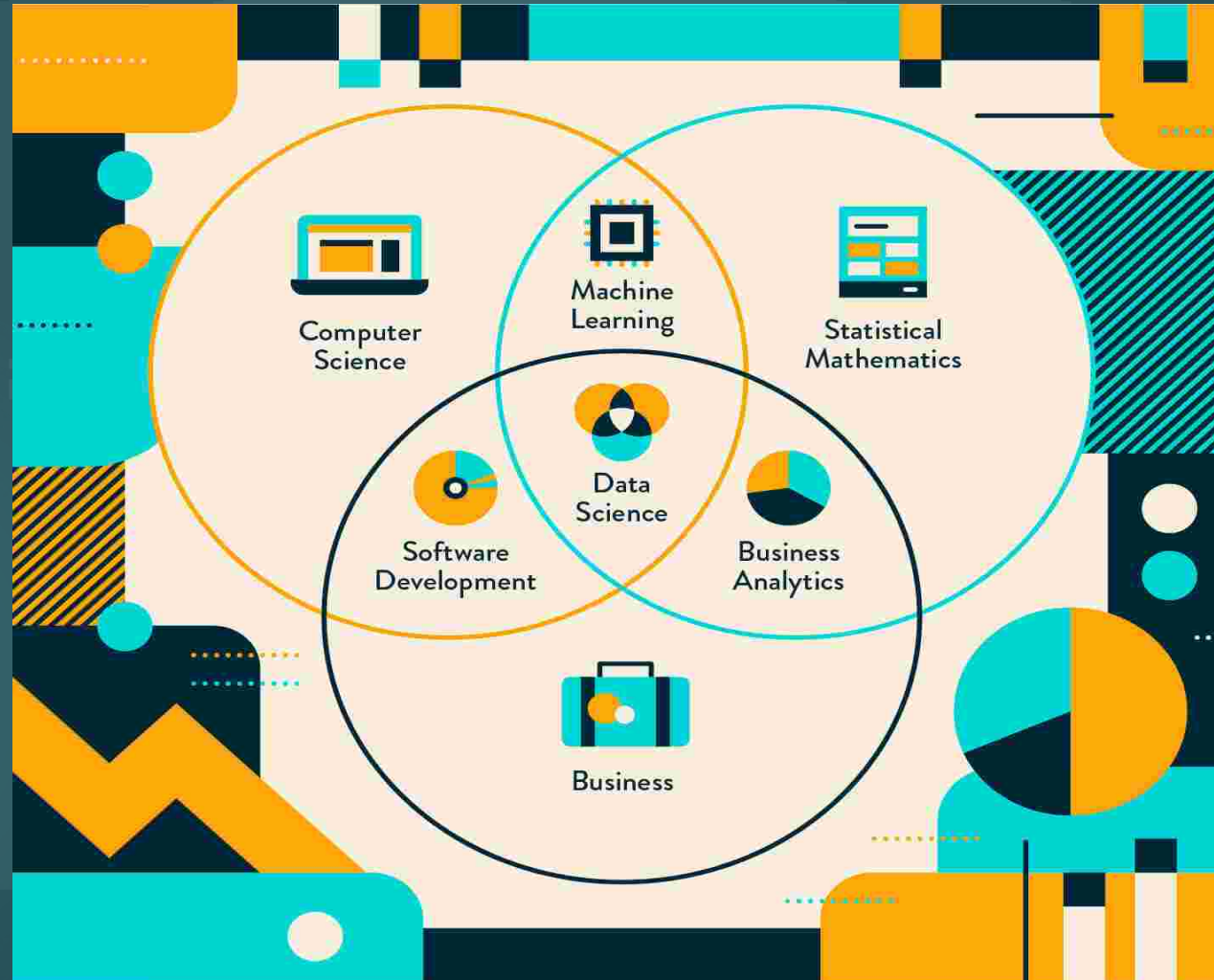
ASMAA MOUNTASSIR, CO-FONDATRICE DAKAE TECH

FATIMA ZAHRA LAHLOU, CO-FONDATRICE DAKAE TECH

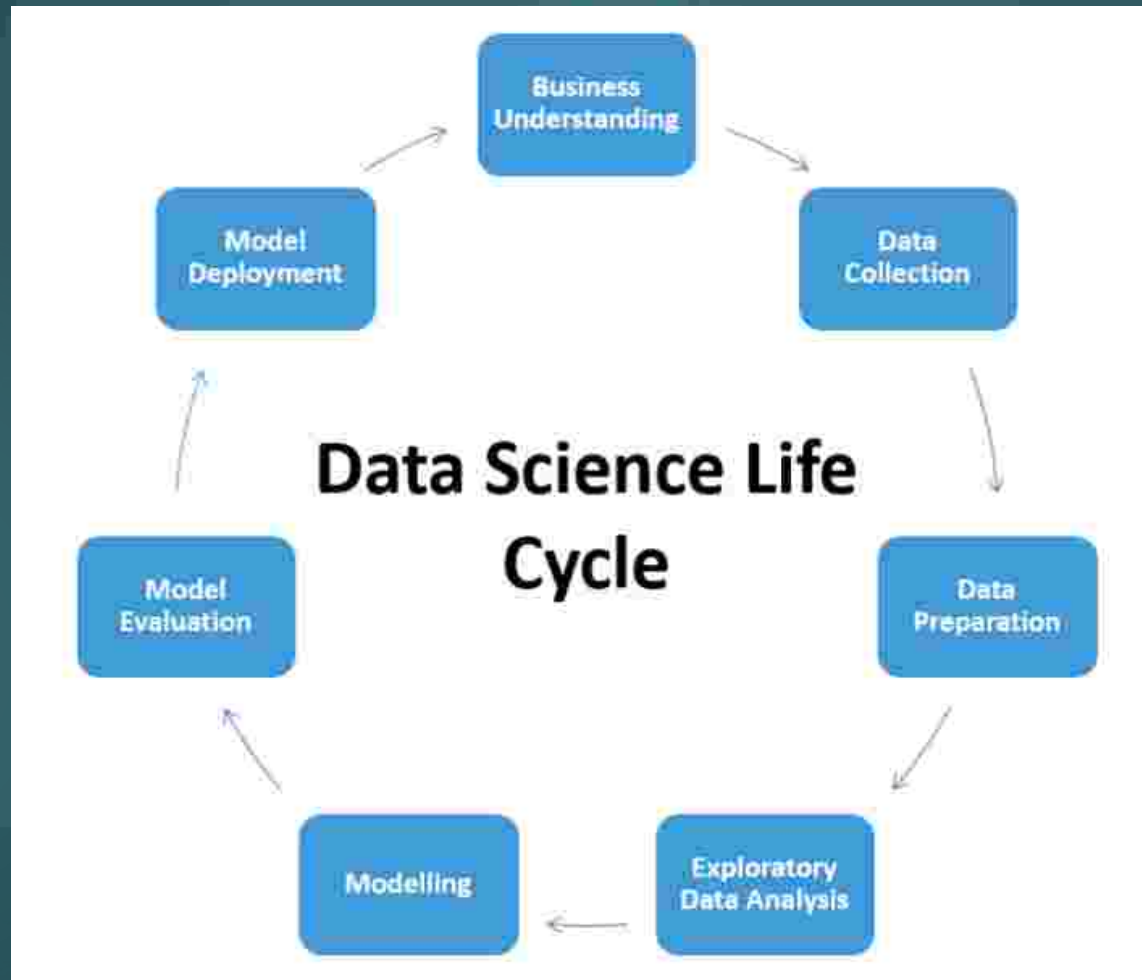
Data Science



Data Science

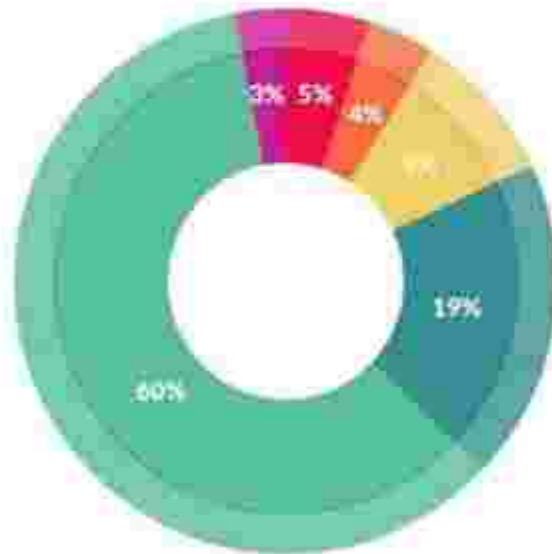


Data Science Project Lifecycle



What is the most time consuming task?

Data preparation accounts for about 80% of the work of data scientists



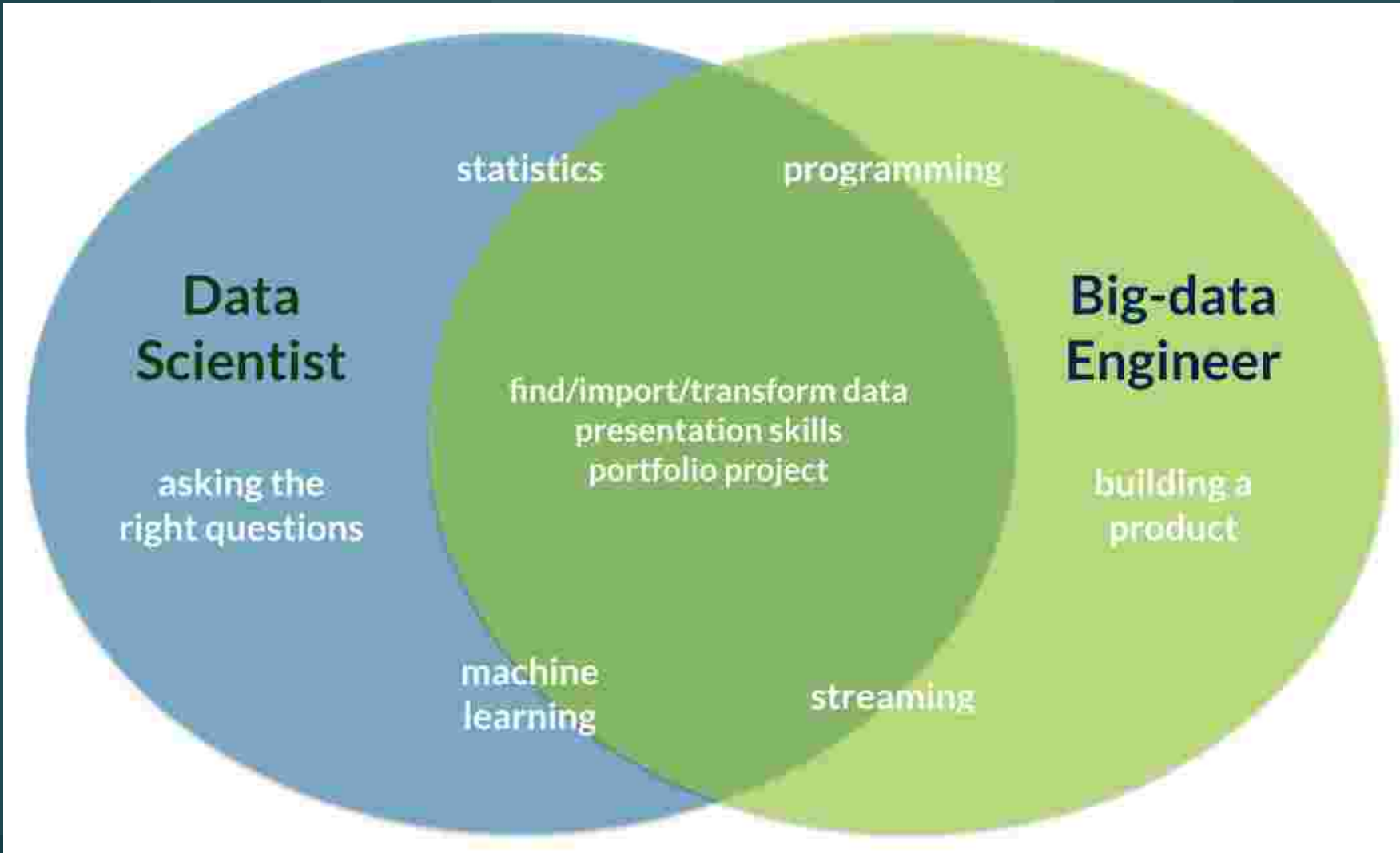
What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Big Data



Big Data



What About my Industry?





- Shelf-space optimization
- Pricing
- Product Bundling
- Promotions
- Up-sell / Cross-sell



- Risk Modeling
- Fraud Prediction
- Customer Segmentation
- Portfolio Optimization



- Market Basket Analytics
- Marketing Mix Analysis
- Product Pricing
- Sales Forecasting
- A/B Testing



- Clinical Trials of New Drugs
- A/B Testing
- Genetics Analysis
- Epidemic Forecasting & Control



- Customer Acquisition Strategies
- Churn Analytics and Control
- Up-Sell/ Cross-Sell
- Product Bundling



- A/B Testing
- Recommendation Engine
- Media Mix Analytics
- Advertisements Optimization
- Customer Acquisition

Analytics is used by almost all industries and functional areas





SEGMENTATION DES CLIENTS

PLAN

▶ Segmentation des Clients

- Quoi ?
- Pourquoi?
- Comment?

▶ Clients de FLY

- Données brutes
- Problématique

▶ Axes du projets

- Dédoublonnage des clients
- Correction des adresses
- Segmentation :
 - Définition des critères
 - Clustering Automatisé



CHAPITRE

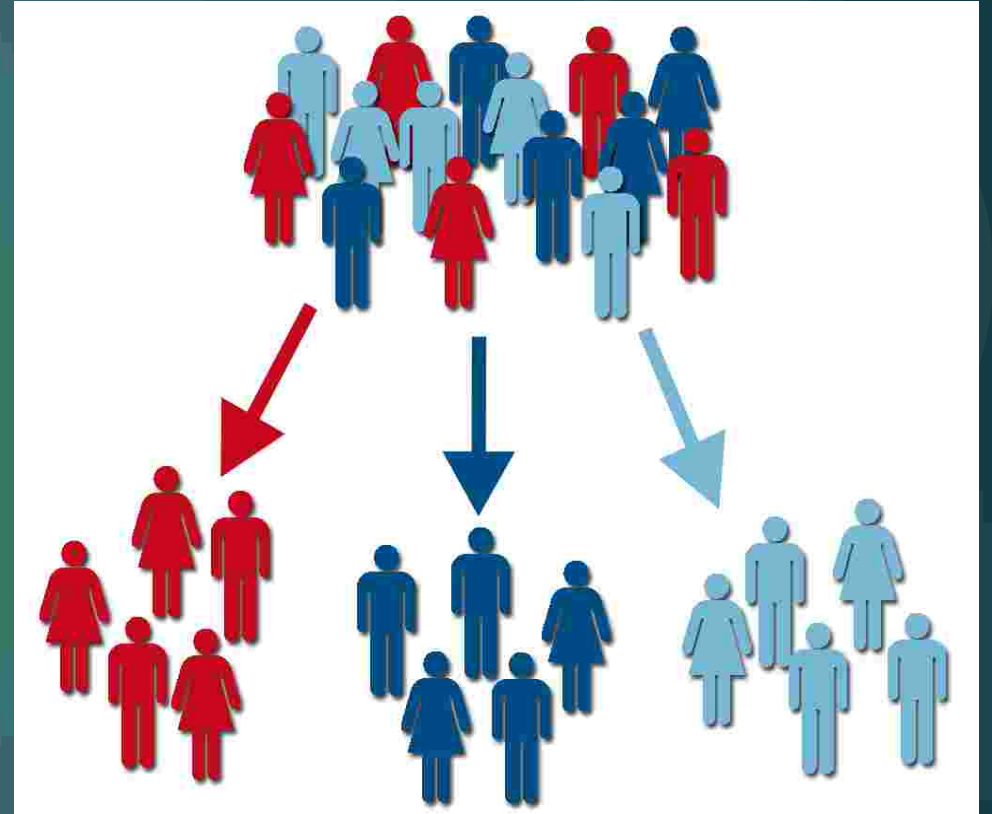
01

SEGMENTATION DES CLIENTS

SEGMENTATION DES CLIENTS

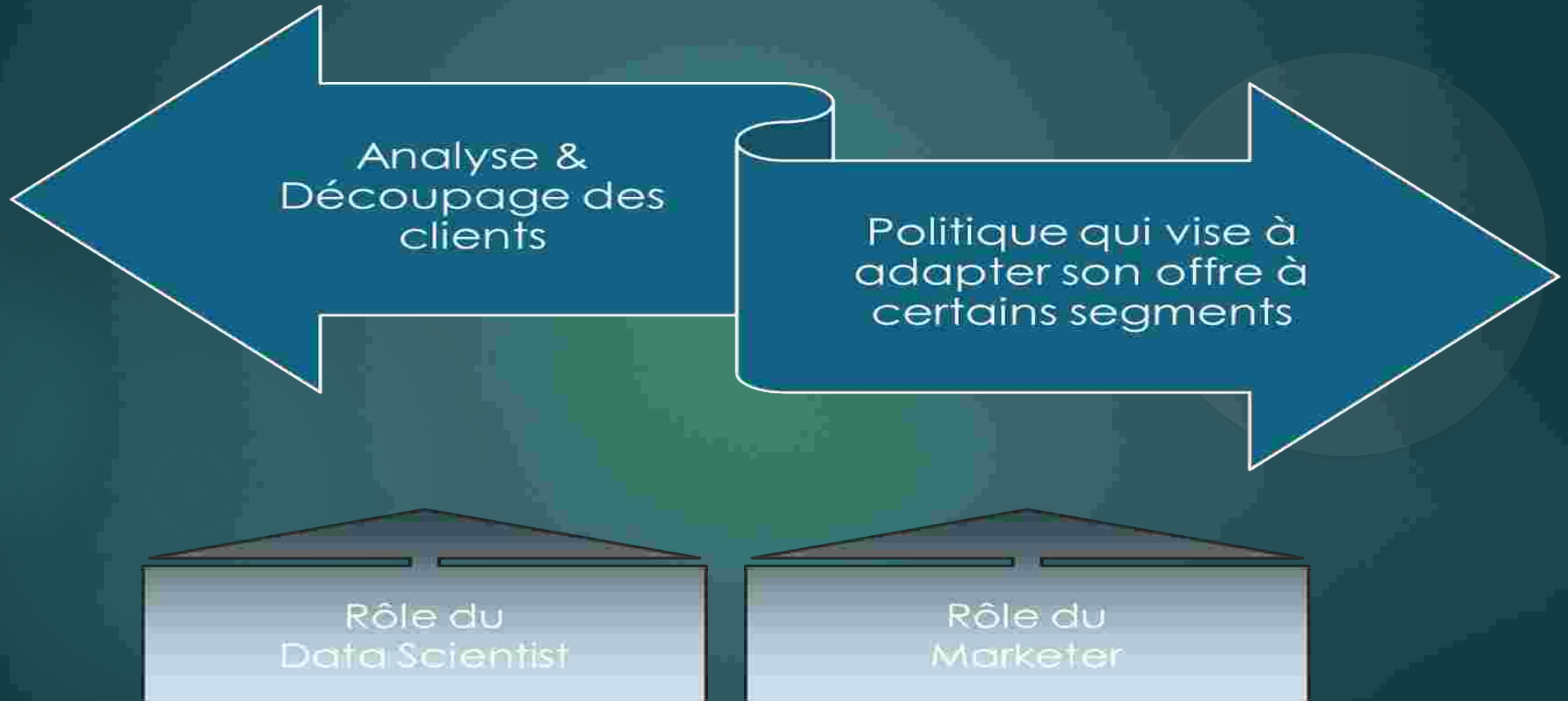
QUOI?

- **Découpage** d'une population (clients, prospects) en **groupes** homogènes
- Différents **critères** (Données socio-démographiques, comportement d'achat, ..).
- Les clients du même segment doivent se ressembler au maximum selon certains critères
- Les clients de segments différents doivent être différents le maximum possible



SEGMENTATION DES CLIENTS

QUOI?



SEGMENTATION DES CLIENTS

POURQUOI?



Marketing de Masse

- Proposer une même offre à l'ensemble du marché
- Coûteux, moins performant

Marketing Ciblé

- Connaître ses clients
- Proposer des offres Adaptées à des segments de clients



SEGMENTATION DES CLIENTS COMMENT?



1. Définir les critères

2. Regrouper les clients en segments selon leur ressemblance

3. Développer des profils types de chaque segment

SEGMENTATION DES CLIENTS

COMMENT?



SEGMENTATION DES CLIENTS

COMMENT?



Segmentation Apriori

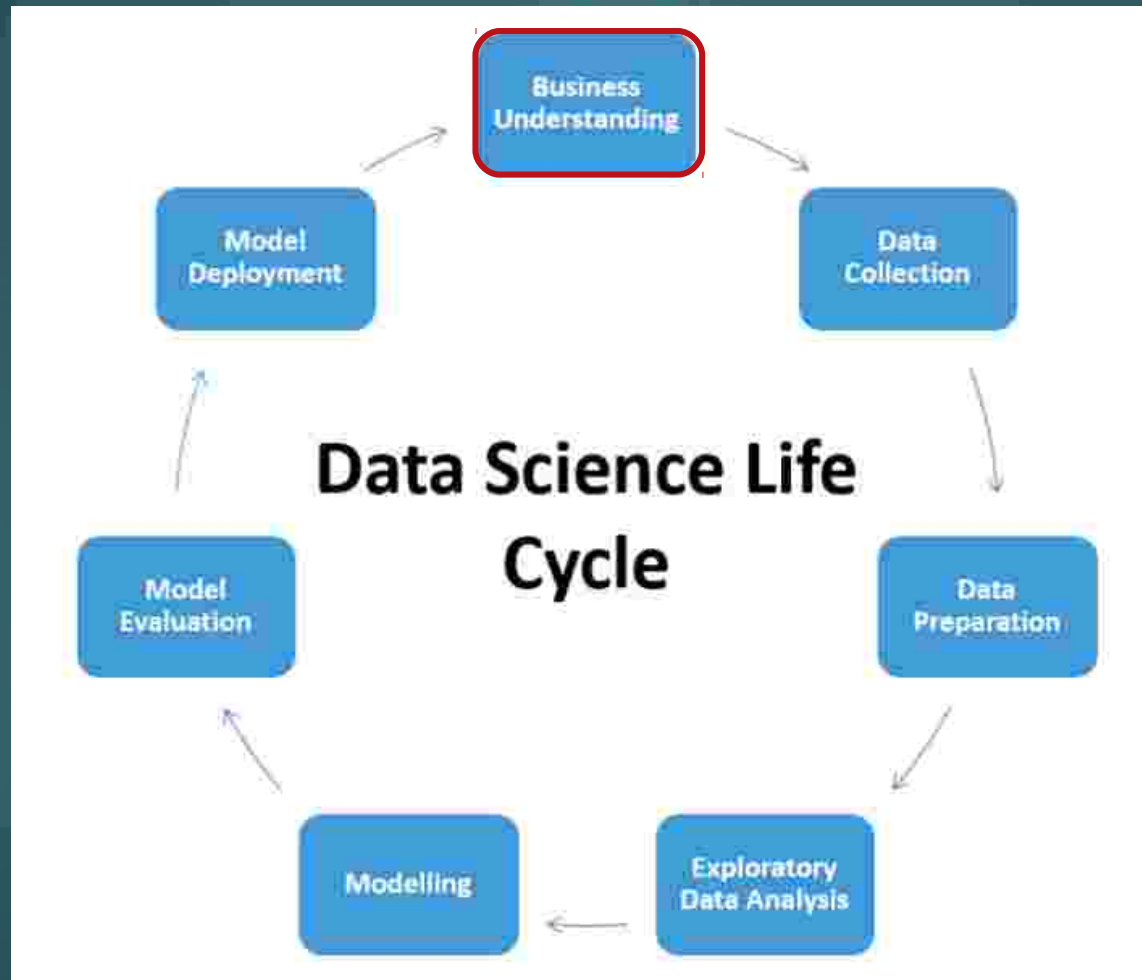
- Modèle basé sur un schéma de variables populaire → pas de particularité !
- Modèle non basé sur les données clients disponibles chez l'entreprise
- Modèle nécessitant des informations personnelles sur les clients (age, revenu, sexe, éducation,...) → non disponibles
- Développement rapide et peu coûteux
- Aperçu faible sur les tendances des clients



Segmentation Post-hoc

- Modèle se basant sur une variété de variables
- Exploitation de la « mine » de données, sur les achats, disponibles chez l'entreprise
- Utilisation de méthodes statistiques et automatiques pour décomposer la clientèle et découvrir les groupes « latents »
- Modélisation dynamique au fil du temps

Data Science Project Lifecycle





CHAPITRE

02

CLIENTS DE FLY



CLIENTS DE FLY DONNÉES BRUTES

TIC	ENT_ID	NomPrenom	Adresse	CMD	CLT_CODE	MAGASIN	DATE_CMD	CODE_RAY	CIV	QTE	MNT	ARTICLE	FAMILLE	SOUS_FAM
	15597319	B ATAILLON FRERES	AV ARISTIDE BRIAND	T345508		V05	27/10/2013	6 SARL		2	626.98	60185002	60	601
	12194941	B ATTITUDE	JEAN LONG 1	T316914		632	12/05/2011	6 ASS		3	297	77706469	69	694
	12205549	B ATTITUDE	JEAN LONG 1	T316951		632	12/05/2011	6 ASS		1	99	77706469	69	694
	12205549	B ATTITUDE	JEAN LONG 1	T316951		632	12/05/2011	7 ASS		2	3.8	77712325	71	718
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		1	3.9	77704079	70	707
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		1	6.9	77714778	70	705
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		1	6.9	77714778	70	705
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		1	12.9	77714403	74	778
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		5	9.5	77712325	71	718
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	18/04/2011	7 ASS		1	15.9	74080090	74	740
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	21/04/2011	6 ASS		1	139	77706470	69	695
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632	21/04/2011	6 ASS		2	79.8	77707599	88	674
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632		ASS		1	179	77702538	69	692
	12219084	B ATTITUDE	JEAN LONG 1	T315676		632		ASS		1	99	77706469	69	694
	14152451	B HOME	1 AV FOCH	983740		892		STE		3	23.7	77723545	75	750
	14152451	B HOME	1 AV FOCH	983740		892		STE		1	14.9	77739980	75	750
	14152451	B HOME	1 AV FOCH	983740		892		STE		2	39.8	77703967	75	751
	14152451	B HOME	1 AV FOCH	983740		892		STE		1	14.9	77739980	75	750
	14152451	B HOME	1 AV FOCH	983740		892		STE		1	12.9	77729569	75	750
	14233711	B HOME	1 AVENUE FOCH	984604		892		STE		1	0.25	77701911	93	833
	14233711	B HOME	1 AVENUE FOCH	984604		892		STE		4	51.6	77729569	75	750
	14233711	B HOME	1 AVENUE FOCH	984604		892		STE		3	44.7	77739980	75	750
	6854449	BERNARD ILLORE	83 RUE DE LA LOUISIE	T184297		V00		MR		1	89	77717833	60	605
	6854449	BERNARD ILLORE	83 RUE DE LA LOUISIE	T184297		V00	15/06/2011	7 MR		2	19.8	77714465	71	726
	19406803	B IMMO	4 RUE LASCARIS	TH01126		794	20/09/2014	6 MR		1	299	77713576	65	650
	19406803	B IMMO	4 RUE LASCARIS	TH01126		794	20/09/2014	6 MR		3	267	77778184	66	665
	17431077	B J M CIE	BRASSERIE DE L'ESCALE	349791		F90	11/03/2014	7 STE		12	23.88	77758135	74	766

Différents
tickets
associés au
même
client !!

CLIENTS DE FLY PROBLÉMATIQUE

Clients
stockés



TIC_ENT_ID	TIC_ENT_CIV	TIC_ENT_NOM_CLI	TIC_ENT_ADR_1	TIC_ENT_ADR_2	TIC_ENT_COMMUNE
10041784	MME	ALONZEAU CARINE	23 BIS RUE LUCIEN BR		PONTAULT COMBAULT
12047281		ALONZEAU CARINE	23-B RUE LUCIEN BRUNET		PONTAULT COMBAULT
16323186	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRAND CHAMPS		BALZAC
16906480	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRANDS CHAMPS		BALZAC
4677171	MME	ALONZO DOLORES	10 RUE LO GRILH		PALAJA
4707831	MME	ALONZO DOLORES	10 LO GRILH		PALAJA
14633755	MME	ALTMAYERHENZIEN EMMANUEL	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
12409378	MME	ALTMAYERHENZIEN EMMANUELLE	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
10122910	Mlle	BONNERONT LIZA	29 31 AVENUE DU CONS	GENERAL NORDLING	LIVRY GARGAN
11914998	MME	BONNERONT LIZA	29 31 AVENUE DU CONSUL	GENERAL NORDLING	LIVRY GARGAN
11746093	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURGUIGNONS		MONTLHERY
10454156	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURG		MONTLHERY
16505302	MME	BONNES ANDRE CATHERINE	MAS ANGLADEROLITE DE CANET		PERPIGNAN
20637393	MME	BONNES ANDRE CATHERINE	MAS ANGLADE ROLITE DE CANET		PERPIGNAN
14714218	MR	BONNESOEUR BENOIST	65 CH DE LA VOUIVRE		MIEUSSY
4964342	MR	BONNESOEUR BENOIT	65 CHEMIN DE LA VOUT		MIEUSSY
7377853	MR	ZANINI JACQUES	32 AVENUE JULES FERRY		THONON LES BAINS
5399420	MR	ZANNINI JACQUES	32 AVENUE JULES FGERRY		THONON LES BAINS

CLIENTS DE FLY PROBLÉMATIQUE

Clients
stockés



TIC_ENT_ID	TIC_ENT_CIV	TIC_ENT_NOM_CLI	TIC_ENT_ADR_1	TIC_ENT_ADR_2	TIC_ENT_COMMUNE
10041784	MME	ALONZEAU CARINE	23 BIS RUE LUCIEN BR		PONTAULT COMBAULT
12047281		ALONZEAU CARINE	23-B RUE LUCIEN BRUNET		PONTAULT COMBAULT
16323186	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRAND CHAMPS		BALZAC
16906480	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRANDS CHAMPS		BALZAC
4677171	MME	ALONZO DOLORES	10 RUE LO GRILH		PALAJA
4707831	MME	ALONZO DOLORES	10 LO GRILH		PALAJA
14633755	MME	ALTMAYERHENZIEN EMMANUEL	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
12409378	MME	ALTMAYERHENZIEN EMMANUELLE	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
10122910	Mlle	BONNERONT LIZA	29 31 AVENUE DU CONS	GENERAL NORDLING	LIVRY GARGAN
11914998	MME	BONNERONT LIZA	29 31 AVENUE DU CONSUL	GENERAL NORDLING	LIVRY GARGAN
11746093	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURGIGNONS		MONTHERY
10454156	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURG		MONTHERY
16505302	MME	BONNES ANDRE CATHERINE	MAS ANGLADEROLITE DE CANET		PERPIGNAN
20637393	MME	BONNES ANDRE CATHERINE	MAS ANGLADE ROLITE DE CANET		PERPIGNAN
14714218	MR	BONNESOEUR BENOIST	65 CH DE LA VOUIVRE		MIEUSSY
4964342	MR	BONNESOEUR BENOIT	65 CHEMIN DE LA VOUT		MIEUSSY
7377853	MR	ZANINI JACQUES	32 AVENUE JULES FERRY		THONON LES BAINS
5399420	MR	ZANNINI JACQUES	32 AVENUE JULES FGERRY		THONON LES BAINS

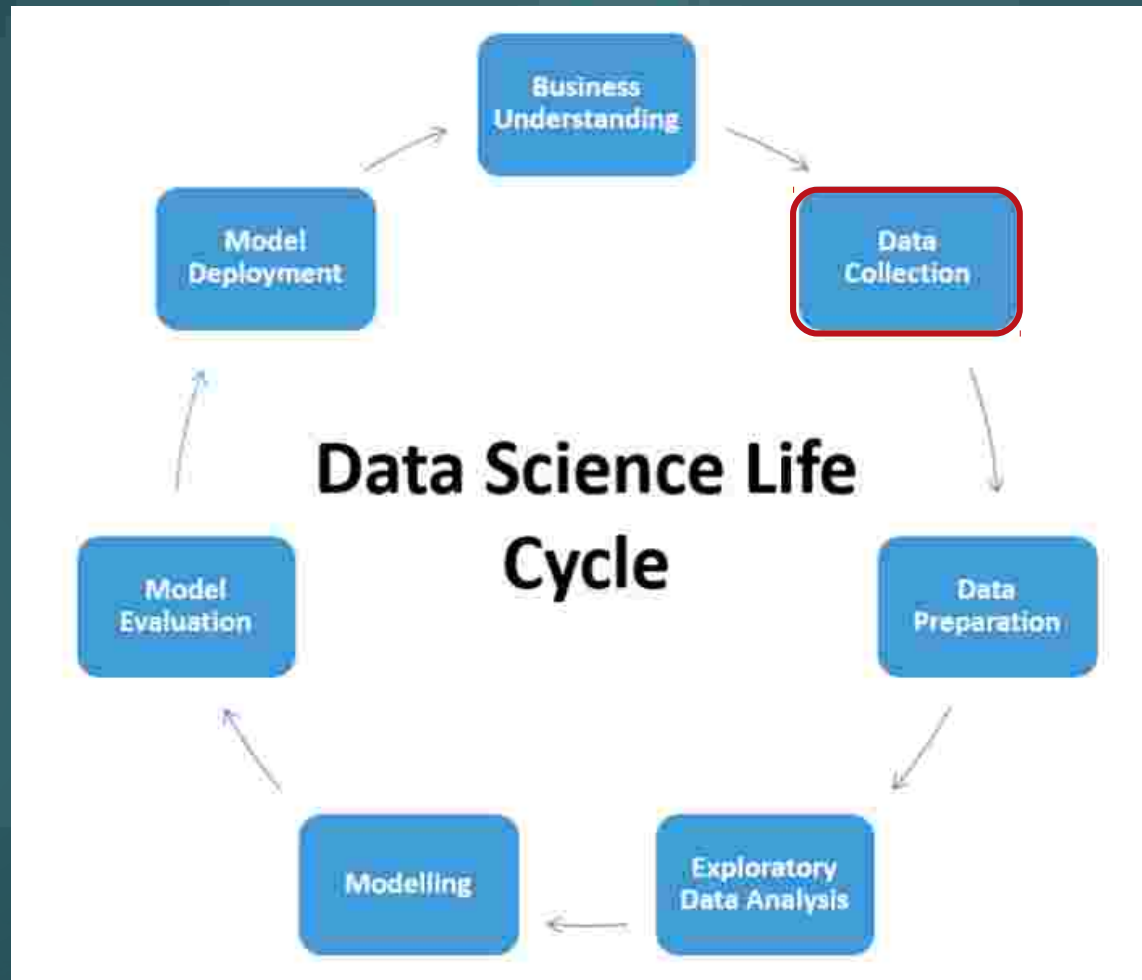
CLIENTS DE FLY PROBLÉMATIQUE

Clients
stockés



TIC_ENT_ID	TIC_ENT_CIV	TIC_ENT_NOM_CLI	TIC_ENT_ADR_1	TIC_ENT_ADR_2	TIC_ENT_COMMUNE
10041784	MME	ALONZEAU CARINE	23 BIS RUE LUCIEN BR		PONTAULT COMBAULT
12047281		ALONZEAU CARINE	23-B RUE LUCIEN BRUNET		PONTAULT COMBAULT
16323186	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRAND CHAMPS		BALZAC
16906480	MME	ALONZO CHRISTELLE	10 IMPASSE DES GRANDS CHAMPS		BALZAC
4677171	MME	ALONZO DOLORES	10 RUE LO GRILH		PALAJA
4707831	MME	ALONZO DOLORES	10 LO GRILH		PALAJA
14633755	MME	ALTMAYERHENZIEN EMMANUEL	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
12409378	MME	ALTMAYERHENZIEN EMMANUELLE	16 RUE DU COMMANDANT BICHERAY		LE HAVRE
10122910	Mlle	BONNERONT LIZA	29 31 AVENUE DU CONS	GENERAL NORDLING	LIVRY GARGAN
11914998	MME	BONNERONT LIZA	29 31 AVENUE DU CONSUL	GENERAL NORDLING	LIVRY GARGAN
11746093	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURGUIGNONS		MONTLHERY
10454156	MR	BONNEROT GERMAIN	10 IMPASSE DES BOURG		MONTLHERY
16505302	MME	BONNES ANDRE CATHERINE	MAS ANGLADEROLITE DE CANET		PERPIGNAN
20637393	MME	BONNES ANDRE CATHERINE	MAS ANGLADE ROLITE DE CANET		PERPIGNAN
14714218	MR	BONNESOEUR BENOIST	65 CH DE LA VOUIVRE		MIEUSSY
4964342	MR	BONNESOEUR BENOIT	65 CHEMIN DE LA VOUT		MIEUSSY
7377853	MR	ZANINI JACQUES	32 AVENUE JULES FERRY		THONON LES BAINS
5399420	MR	ZANNINI JACQUES	32 AVENUE JULES FGERRY		THONON LES BAINS

Data Science Project Lifecycle



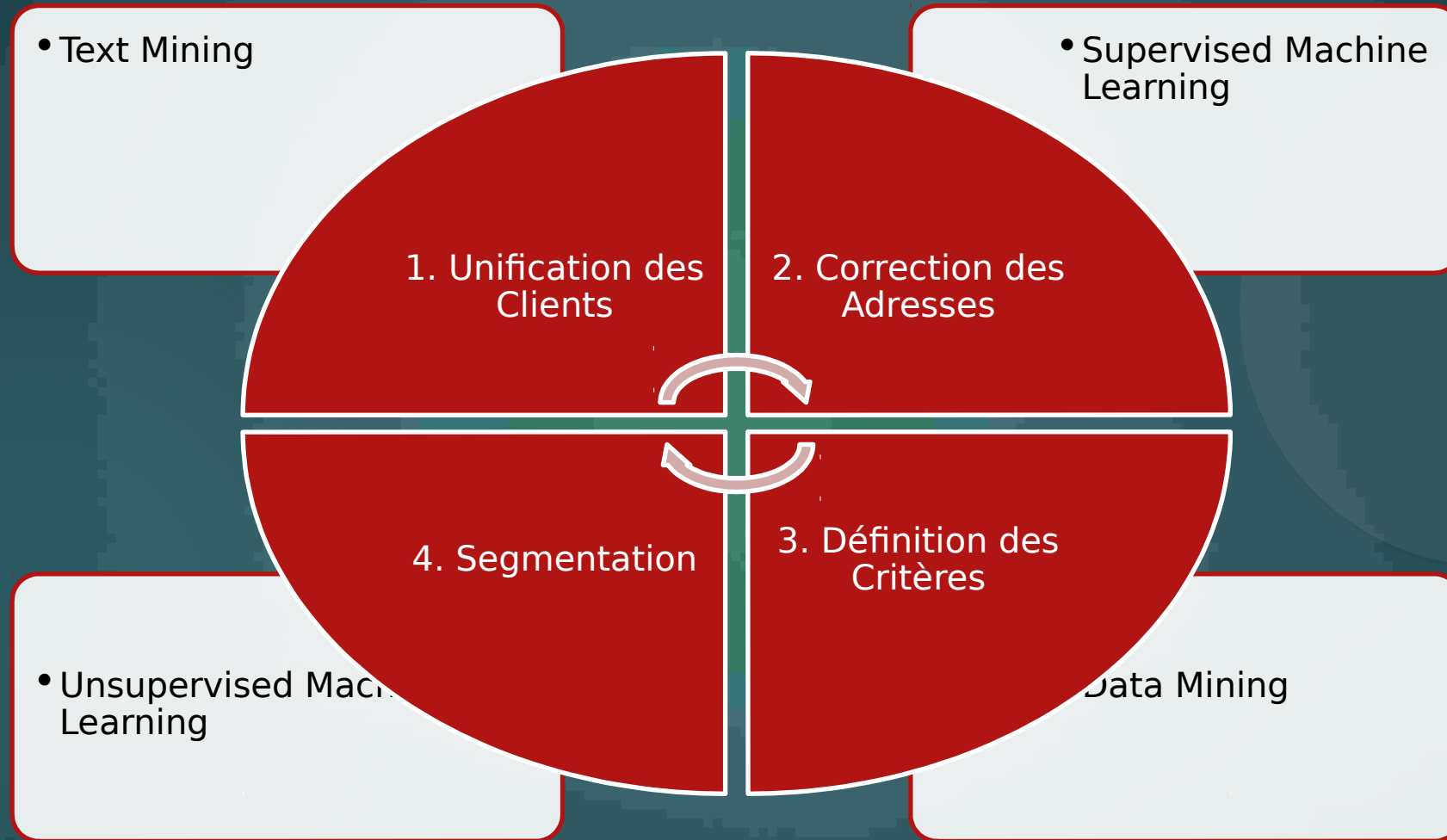


CHAPITRE

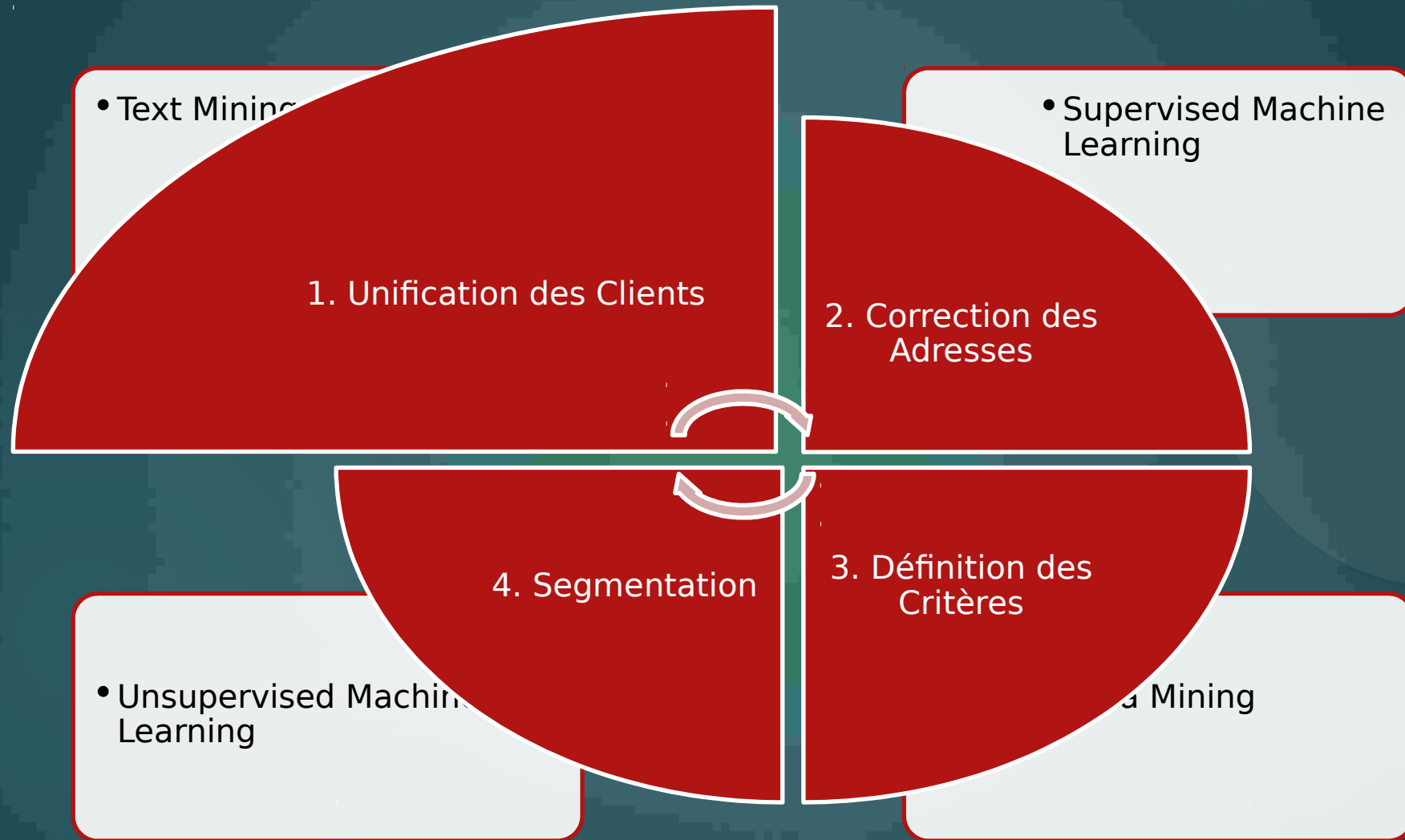
03

AXES DU PROJET

AXES DU PROJETS



AXES DU PROJETS



AXES DU PROJETS

UNIFICATION DES CLIENTS

Charger tous les
clients

Ignorer les clients
inconnus

Appliquer une
technique de
Text Mining pour
mesurer les
similarités entre
les coordonnées

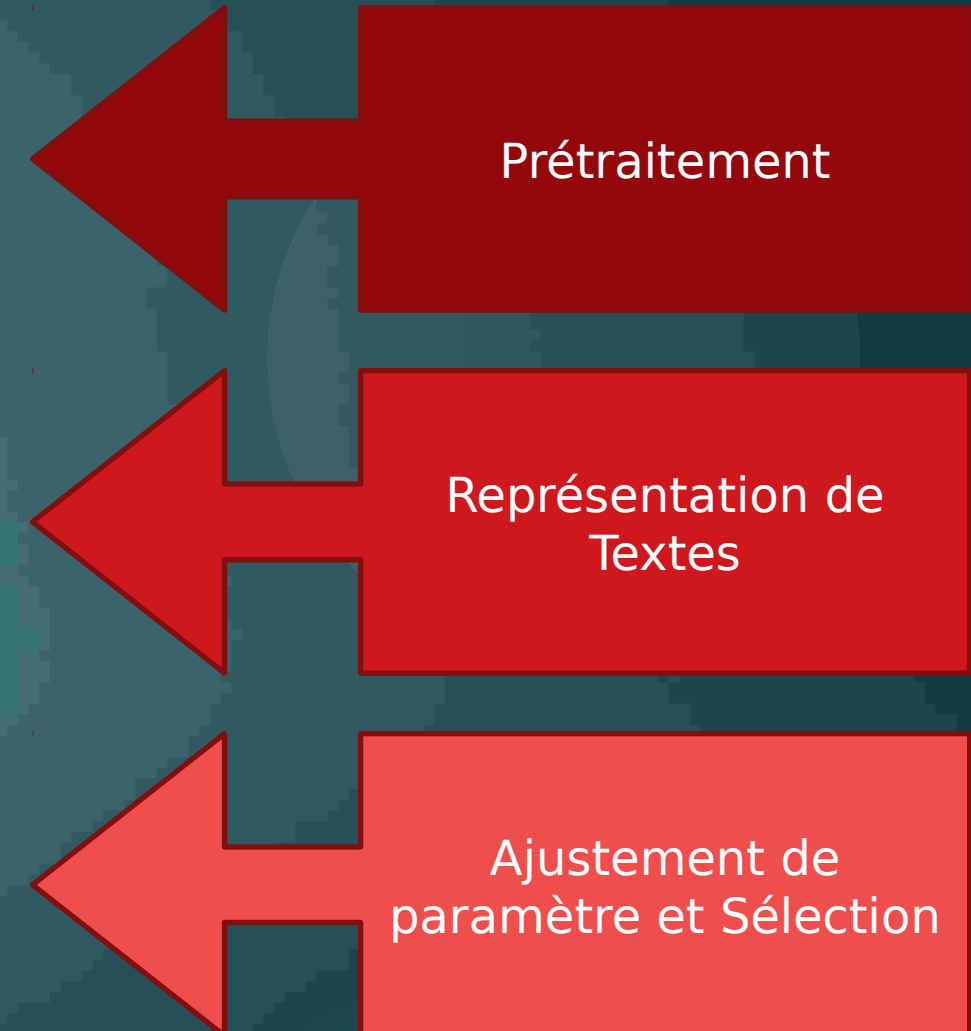
Identifier les
coordonnées les
plus similaires
(similarité > 54%)

AXES DU PROJETS

UNIFICATION DES CLIENTS

1. Concaténer les éléments de coordonnées : Nom + Prénom + Civilité + Adresse1 + Adresse2 + Ville
2. Enlever l'espace, ponctuation, minuscule,...
3. Associer à chaque client une chaîne de caractère : coordonnée
4. Représenter l'ensemble des coordonnées : modèle n-grammes (n=4) et pondération TFIDF
5. Multiplier la matrice obtenue par sa transposée
6. Pour chaque score de la matrice : $x(i,j) \leftarrow (x(i,j) * 100) / x(i,i)$
7. Si $\text{score}(i,j) > 54$: les coordonnées i et j représentent le même client

Remarque : le seuil 54 a été ajusté sur des jeux de données de test



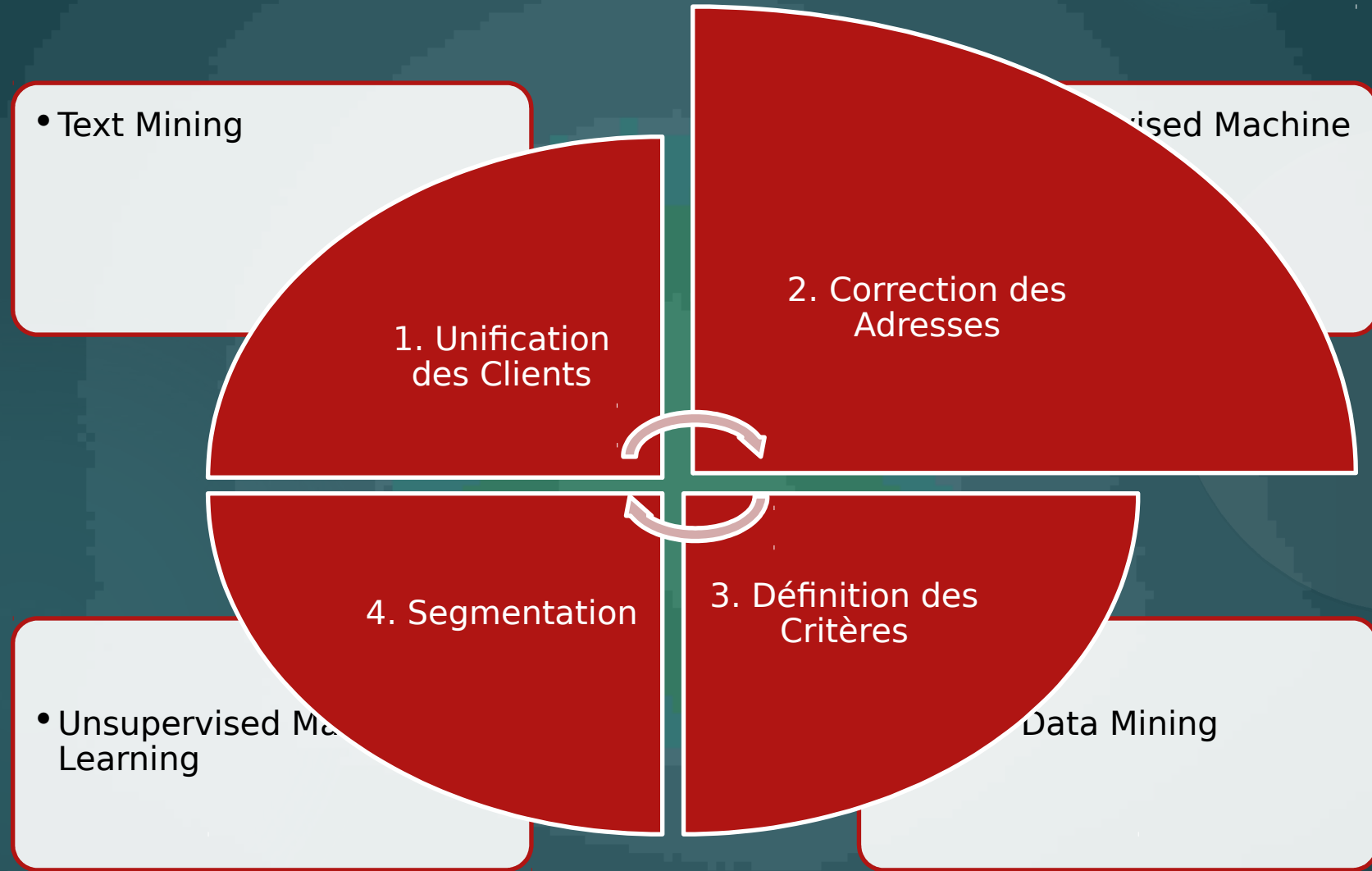
AXES DU PROJET

UNIFICATION DES CLIENTS

TIC_ENT_ID	COORDCLT
7377853	ZANNINI JACQUES MR 32 AVENUE JULES FERRY THIONVILLE LES BAINS
5399420	ZANNINI JACQUES MR 32 AVENUE JULES FGERRY THIONVILLE LES BAINS
10041784	ALONZEAU CARINE MME 23 BIS RUE LUCIEN BR. PONTAULT COMBAULT
12047281	ALONZEAU CARINE 23 B RUE LUCIEN BRUNET PONTAULT COMBAULT
4677171	ALONZO DOLORES MME 10 RUE LO GRILH PALAJA
4707831	ALONZO DOLORES MME 10 LO GRILH PALAJA
16323186	ALONZO CHRISTELLE MME 10 IMPASSE DES GRAND CHAMPS BALZAC
16906480	ALONZO CHRISTELLE MME 10 IMPASSE DES GRANDS CHAMPS BALZAC

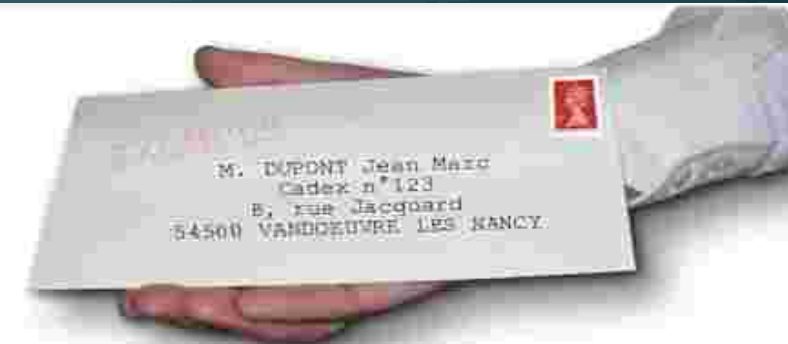


AXES DU PROJETS



AXES DU PROJET

CORRECTION DES ADRESSES



AXES DU PROJET

CORRECTION DES ADRESSES

adresse	cp	commune	newAddress
4 RUE BERTRAND RUSSEL	25000	BESANCON	4 Rue Bertrand Russel 25000 Besançon, France
144 FAUGBOURG BANNIER	45000	ORLEANS	faugbourg bannier 144 Faubourg Bannier 45000 Orléans, France
CHEMIN DE LA VERNIQUE	69130	ECULLY	8 chemin de la vernique 69130 ecully france
58 BIS AVENUE GAMBETTA	94100	ST MAUR DES FOSSES	58bis avenue gambetta 94100 saint-maur-des-fosses france
26 AV PAUL DOUMER	6190	ROQUEBRUNE CAP MARTIN	26 avenue paul doumer 06190 roquebrune-cap-martin france
RUE DU DOCTEUR ROUX MME MEUNIER	49300	CHOLET	20 rue du champ noir 49300 cholet france
280 RUE RICHETTA	69400	VILLEFRANCHE SUR SAONE	280 rue alexandre richetta 69400 villefranche-sur-saone france
ZAC VAL DE DURANCE CH LES TROIS FRANCS	13860	PEYROLLES EN PROVENCE	zone artisanale et commerciale vallee de durance chemin des 3 francs
28 RUE DES OLIVETTES	44000	NANTES	28 rue des olivettes 44000 nantes france
827 AV DE LA VICTOIRE DU 8 MAI	83000	TOULON	827 avenue de la victoire du 8 mai 1945 83000 toulon france
ECOLE DE NEUVECELLE	74500	NEUVECELLE	ecole de neuvecelle Avenue de Maraiche 74500 Neuvecelle, France
1 AVENUE DU CHATEAU HEBERT	35800	DINARD	1 avenue du chateau hebert 35800 dinard france
20 PLACE GABRIEL FAURE	74940	ANNECY LE VIEUX	20 place gabriel faure 74940 annecy-le-vieux france
18 RUE ROCHEVALIER	48100	MARVEJOLS	18 Rue Rochevalier 48100 Marvejols, France
13 CHEMIN DE L'INDUSTRIE	69570	DARDILLY	13 chemin de l'industrie 69570 dardilly france
70 AV DES PEINTRES ROUX	13012	MARSEILLE	70 avenue des peintres roux 13012 marseille france
33 RUE ARSON	6300	NICE	33 rue arson 06300 nice france
AV DES MAUMETTES	6270	VILLENEUVE LOUBET	62 avenue des baumettes 06270 villeneuve-loubet france
27 RUE DARSONVAL	44000	NANTES	27 Rue d'Arsonval 44000 Nantes, France
1 RUE DE LA POISSONNERIE	6300	NICE	1 Rue de la Poissonnerie 06300 Nice, France
MAISON RELAIS 18 RUE DU 22 NOVEMBRE	67000	STRASBOURG	maison relais 18 rue du vingt-deux novembre 67000 strasbourg france

AXES DU PROJET

CORRECTION DES ADRESSES

adresse	cp	commune	newAddress	label
4 RUE BERTRAND RUSSEL	25000	BESANCON	4 Rue Bertrand Russel 25000 Besançon, France	OK
144 FAUGBOURG BANNIER	45000	ORLEANS	faugbourg banner 144 Faubourg Bannier 45000 Orléans, France	OK
CHEMIN DE LA VERNIQUE	69130	ECULLY	8 chemin vernique 69130-ecully.france	OK
58 BIS AVENUE GAMBETTA	94100	ST MAUR DES FO	58bis Gambetta 94100 saint-maur-des-fosses france	OK
26 AV PAUL DOUMER	81000	ROQUEBRUNE	26 avenue paul doumer 81000 roquebrune-cap-martin france	OK
RUE DU DOCTEUR ROUX MME MEUNIER	85000	CHOLET	rue du docteur roux mme meunier 85000 cholet france	NOK
280 RUE RICHETTA	71000	SAONE	280 rue richetta 71000 saone france	OK
ZAC VAL DE DURANCE CH LES	83000	TULON	zac val de durance chemin des 3 francs 83000 toulon france	OK
28 RUE DES OLIVIETTES	74500	NEUVECELLE	28 rue des oliviettes 74500 Neuvecelle, France	NOK
827 AV DE LA VICTOIRE DU 8 MAI	13000	DINARD	827 av de la victoire du 8 mai 13000 dinard france	OK
ECOLE DE NEUVECELLE	74500	NEUVECELLE	ecole de Maraiche 74500 Neuvecelle, France	NOK
1 AVENUE DU CHATEAU HEB	74500	NEUVECELLE	1 avenue du chateau heb 74500 Neuvecelle, France	NOK
20 PLACE GABRIEL FAURE	74500	NEUVECELLE	20 place gabriel faure 74500 Neuvecelle, France	NOK
18 RUE ROCHEVALIER	48100	MARVEJOIS	18 rue rochevalier 48100 Marvejols, France	OK
13 CHEMIN DE L'INDUSTRIE	69570	DARDILLY	13 chemin de l'industrie 69570 dardilly france	OK
70 AV DES PEINTRES ROUX	13012	MARSEILLE	70 avenue des peintres roux 13012 marseille france	OK
33 RUE ARSON	06300	NICE	33 rue arson 06300 nice france	OK
AV DES MAUMETTES	06270	VILLENEUVE LOUBET	62 avenue des baumettes 06270 villeneuve-loubet france	OK
27 RUE DARSONVAL	44000	NANTES	27 Rue d'Arsonval 44000 Nantes, France	OK
1 RUE DE LA POISSONNERIE	06300	NICE	1 Rue de la Poissonnerie 06300 Nice, France	OK
MAISON RELAIS 18 RUE DU 22 NOVEMBRE	67000	STRASBOURG	maison relais 18 rue du vingt-deux novembre 67000 strasbourg france	OK

**8 millions
d'adresses !!!**

AXES DU PROJET

CORRECTION DES ADRESSES

adresse	cp	commune	NLS	JCD	label
4 RUE BERTRAND RUSSEL	25000	BESANCON	Russel 25000 Besancon, France	1	1 OK
144 FAUGBOURG BANNIER	45000	ORLEANS	faubourg Bannier 144 Faubourg Bannier 45000 Orléans, France	0.66037736	0.68888889 OK
CHEMIN DE LA VERNIQUE	69130	ECULLY	8 chemin Vernique 69130 ecully france	1	1 OK
58 BIS AVENUE GAMBETTA	94100	ST MAUR DES FOSSES	58bis avenue Gambetta 94100 saint-maur-des-fosses france	1	1 OK
26 AV PAUL DOUMER	6190	ROQUEBRUNE CAP MARTIN	26 avenue Paul Doumer 06190 roquebrune-cap-martin france	1	1 OK
RUE DU DOCTEUR ROUX MME MEUNIER	49300	CHOLET	20 rue du Docteur Roux MME Meunier 49300 cholet france	0.34482759	0.05714286 NOK
280 RUE RICHETTA	69400	VILLEFRANCHE SUR SAONE	280 rue Richetta 69400 villefranche-sur-saone france	0.55	0.31578947 OK
ZAC VAL DE DURANCE CH LES TROIS FRANCS	13860	PEYROLLES EN PROVENCE	zone d'activite et commerciale vallee de durance chemin des 3 francs	0.9375	0.52941177 OK
28 RUE DES OLIVETTES	44000	NANTES	28 rue des Olivettes 44000 nantes france	0.9375	0.5625 OK
827 AV DE LA VICTOIRE DU 8 MAI	83000	TOULON	827 avenue de la victoire du 8 mai 1945 83000 toulon france	0.85714286	0.84 OK
ECOLE DE NEUVECELLE	74500	NEUVECELLE	ecole neuvecelle Avenue de Maralche 74500 Neuvecelle, France	0.70909091	0.57446809 NOK
1 AVENUE DU CHATEAU HEBERT	35800	DINARD	1 avenue du chateau hebert 35800 dinard france	1	1 OK
20 PLACE GABRIEL FAURE	74940	ANNECY LE VIEUX	20 place gabriel faure 74940 annecy-le-vieux france	0.93333333	0.62745098 OK
18 RUE ROCHEVALIER	48100	MARVEJOLS	18 rue Rochevalier 48100 Marvejols, France	1	1 OK
13 CHEMIN DE L'INDUSTRIE	69570	DARDILLY	13 chemin de l'industrie 69570 dardilly france	0.94736842	0.63157895 OK
70 AV DES PEINTRES ROUX	13012	MARSEILLE	70 avenue des peintres roux 13012 marseille france	1	1 OK
33 RUE ARSON	6300	NICE	33 rue arson 06300 nice france	1	1 OK
AV DES MAUMETTES	6270	VILLENEUVE LOUBET	62 avenue des baumettes 06270 villeneuve-loubet france	0.94444444	0.57894737 OK
27 RUE DARSONVAL	44000	NANTES	27 Rue d'Arsonval 44000 Nantes, France	0.92307692	0.46153845 OK
1 RUE DE LA POISSONNERIE	6300	NICE	1 rue de la Poissonnerie 06300 Nice, France	1	1 OK
MAISON RELAIS 18 RUE DU 22 NOVEMBRE	67000	STRASBOURG	maison relais 18 rue du vingt-deux novembre 67000 strasbourg france	0.56521739	0.28 OK

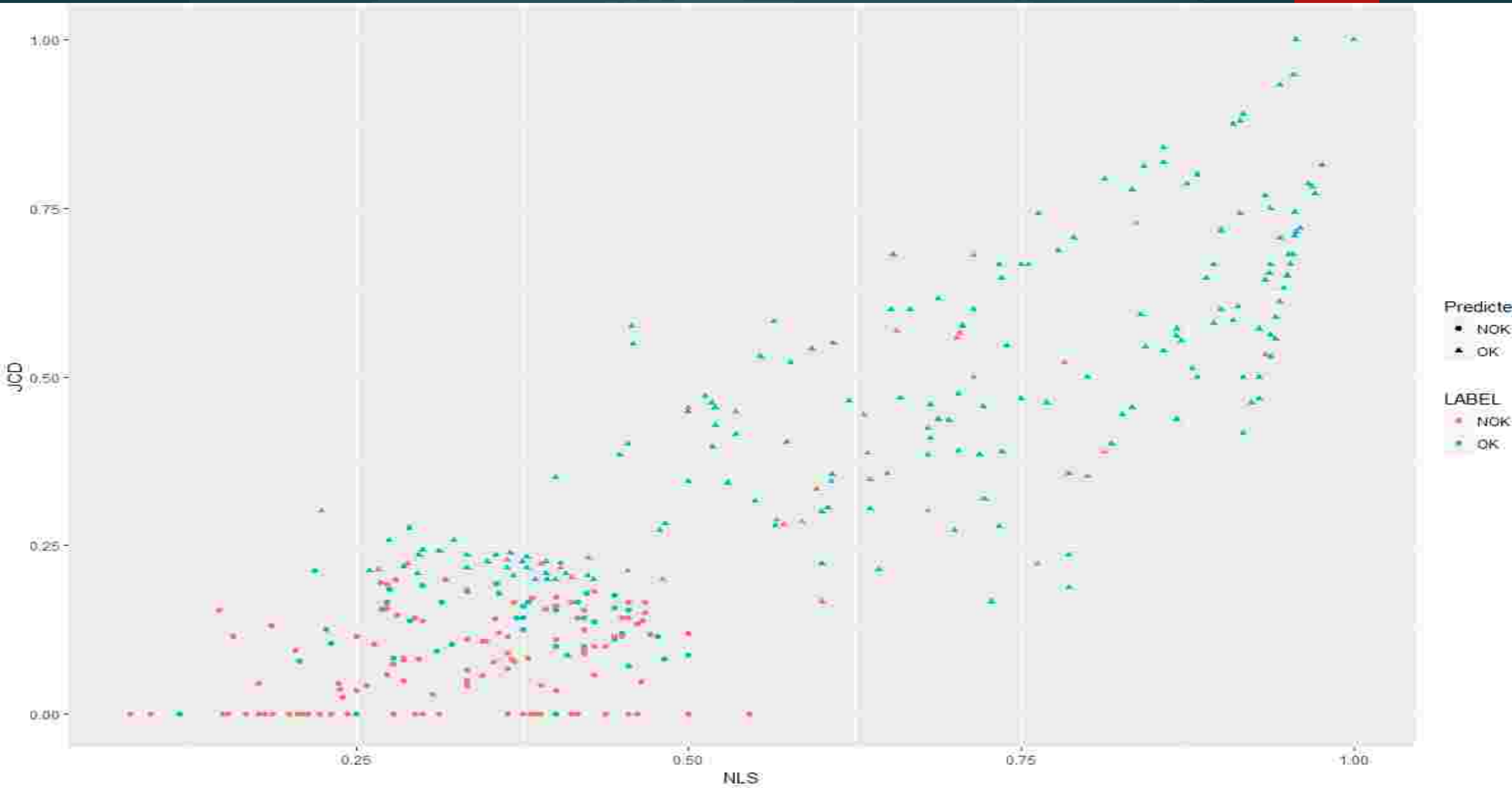
Mesures de similarité

AXES DU PROJET

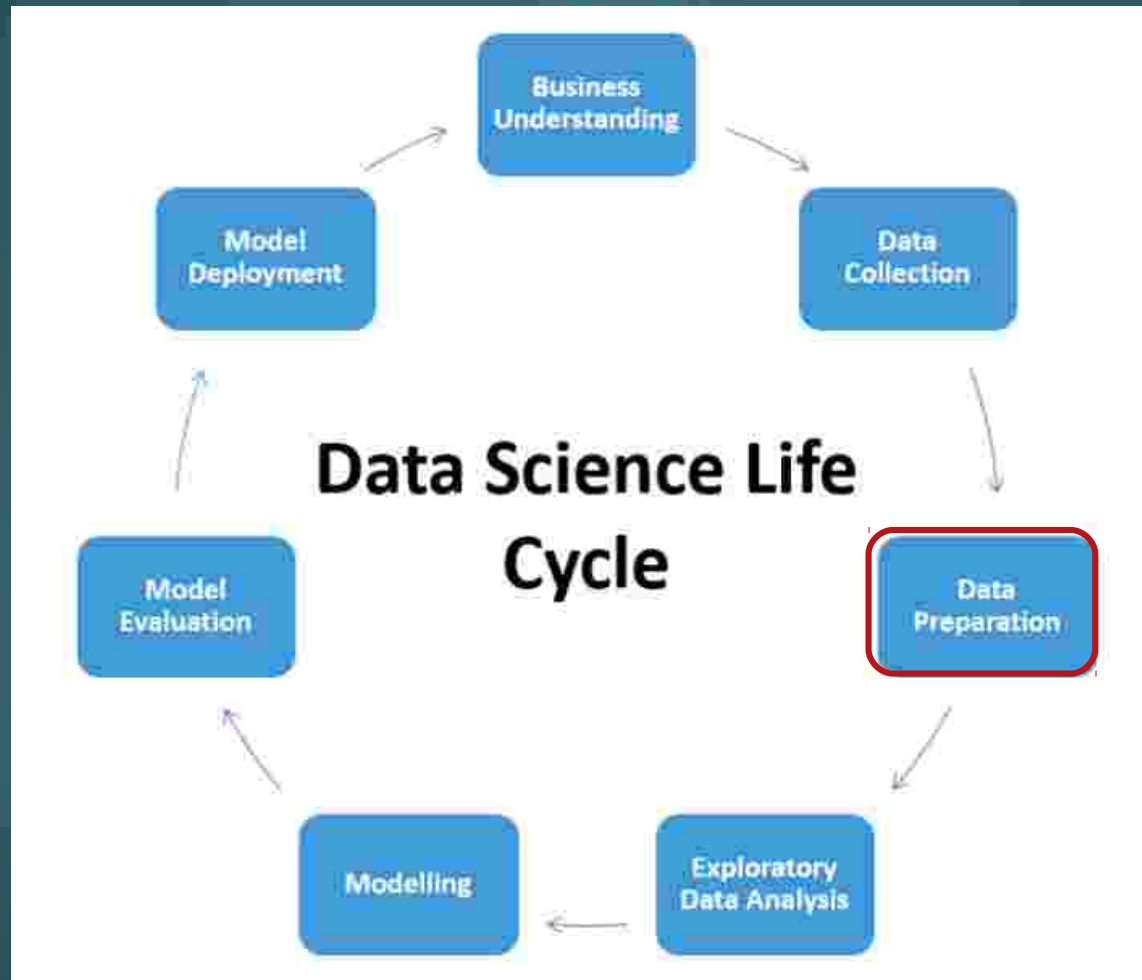
CORRECTION DES ADRESSES



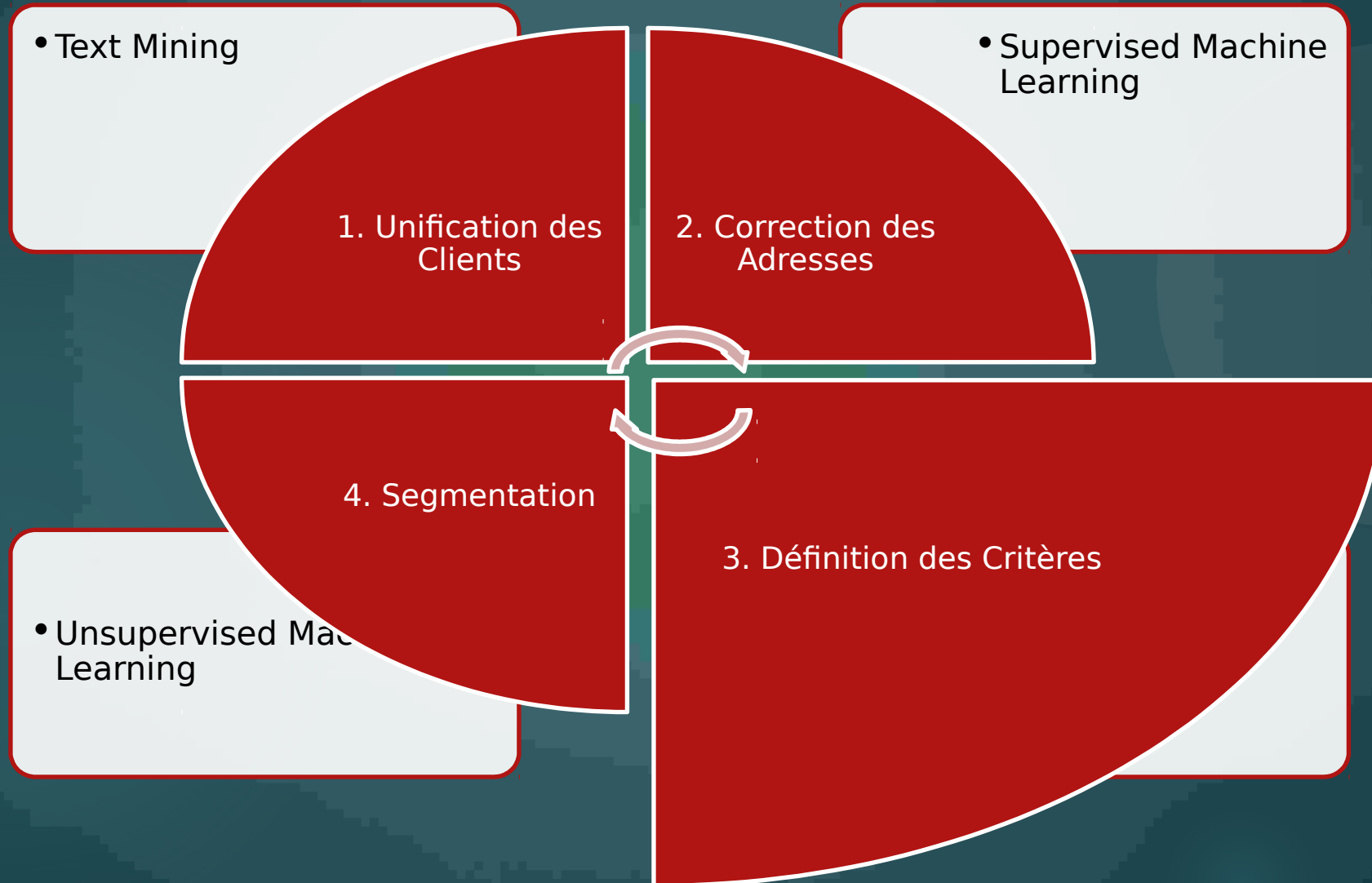
Machine Learning :
SVM !!



Data Science Project Lifecycle



AXES DU PROJETS



AXES DU PROJET

DÉFINITION DES CRITÈRES (1/3)

	Critère	Signification
CLIENT	TicEntId	Id du ticket
	CoordClt	Coordonnées du client
	Magasin	Code du magasin où le client fait la plupart de ses achats. La comparaison entre les magasins se fait à base du nombre de passages
	Civilite	Civilité du client : Mr/Mme/Mlle/Couple/Ste/Inconnu
	Region	Région correspondant à CP_ID
	Porteur	Porteur d'une carte FID (o/n)
	Age	Age du client s'il est porteur
	NewsLet	Abonné à Newsletter (o/n)
	PremierPassage	Date du premier passage
	DernierPassage	Date du passage le plus récent
	Récence	Nombre de jours depuis le dernier passage
	Tenure	Nombre de jours entre le premier passage et le dernier passage
	NbPassages	Nombre de dates marqués sur les tickets du clients

AXES DU PROJET

DÉFINITION DES CRITÈRES (2/3)



	Critère	Signification
ACHATS	MontWeb	Montant des achats effectués via web
	MontPub	Montant d'achat d'articles en PUB
	MontMeuble	Montant d'achat d'articles du rayon Meuble
	MontBoutique	Montant d'achat d'articles du rayon Boutique
	MontTopSelling	Montant d'achat d'articles TopSelling (topSelling pour chaque famille de chaque trimestre) correspondant à la date d'achat
	MontPresta	Montant des prestations de montage et de livraison
	TotalMont	Montant total des achats effectués
	NbWeb	Quantité des articles achetés via web
	NbPub	Quantité des articles en PUB achetés
	NbMeuble	Quantité des articles du rayon Meuble
	NbBoutique	Quantité des articles du rayon Boutique
	NbTopSelling	Quantité des articles TopSelling
	NbTotal	Quantité totale des articles achetés
	HorsWknd	Nombre de passages hors week-end
	NbRetours	Quantité d'articles en retour
	DiversitéProduits	Différentes sous-familles achetées
	MargeMeuble	Marge en articles Meuble achetés
	MargeBoutique	Marge en articles Boutique achetés
	MargeTopSelling	Marge en articles TopSelling achetés
	MargePresta	Marge en prestations de montage et de livraison
	MargeTotal	Marge totale des achats effectués

AXES DU PROJET

DÉFINITION DES CRITÈRES (3/3)

	Critère	Signification
FINANCEMENT	CarteFid	Montant du ticket où une carte FID a été présentée
	CarteFidPlus	Montant du ticket où une carte CTLM a été présentée
	CartePersonnel	Montant du ticket où une carte personnel Fly a été présentée
	PaiPersonnel	Montant payé via financement personnel
	PaiCredit	Montant payé via un crédit
	PaiBons	Montant payé via les bons d'achat,..
	Remise	Montant de remise
	F022	Montant payé par F022
	F023	Montant payé par F023
	TotalPaielement	Montant total payé via les différents moyens

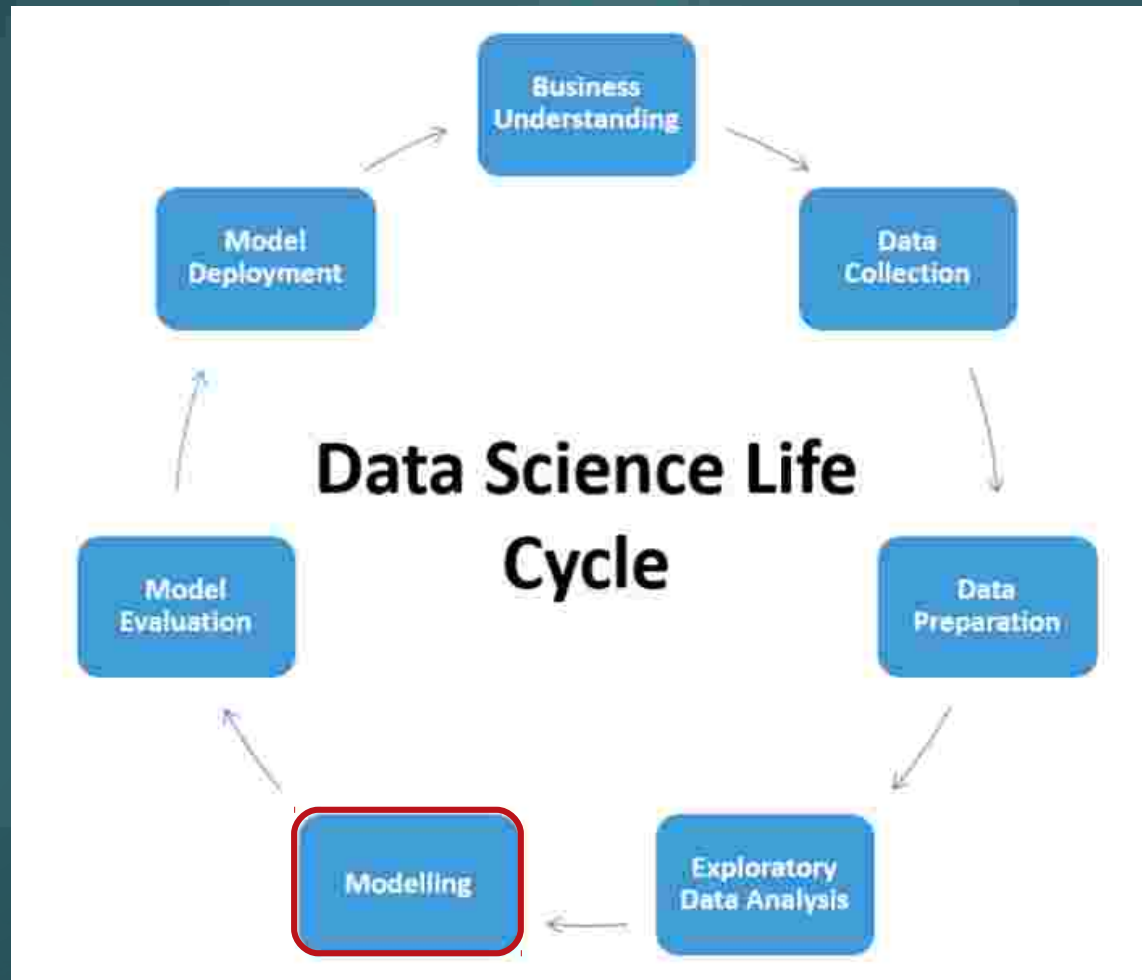
AXES DU PROJET

DÉFINITION DES CRITÈRES

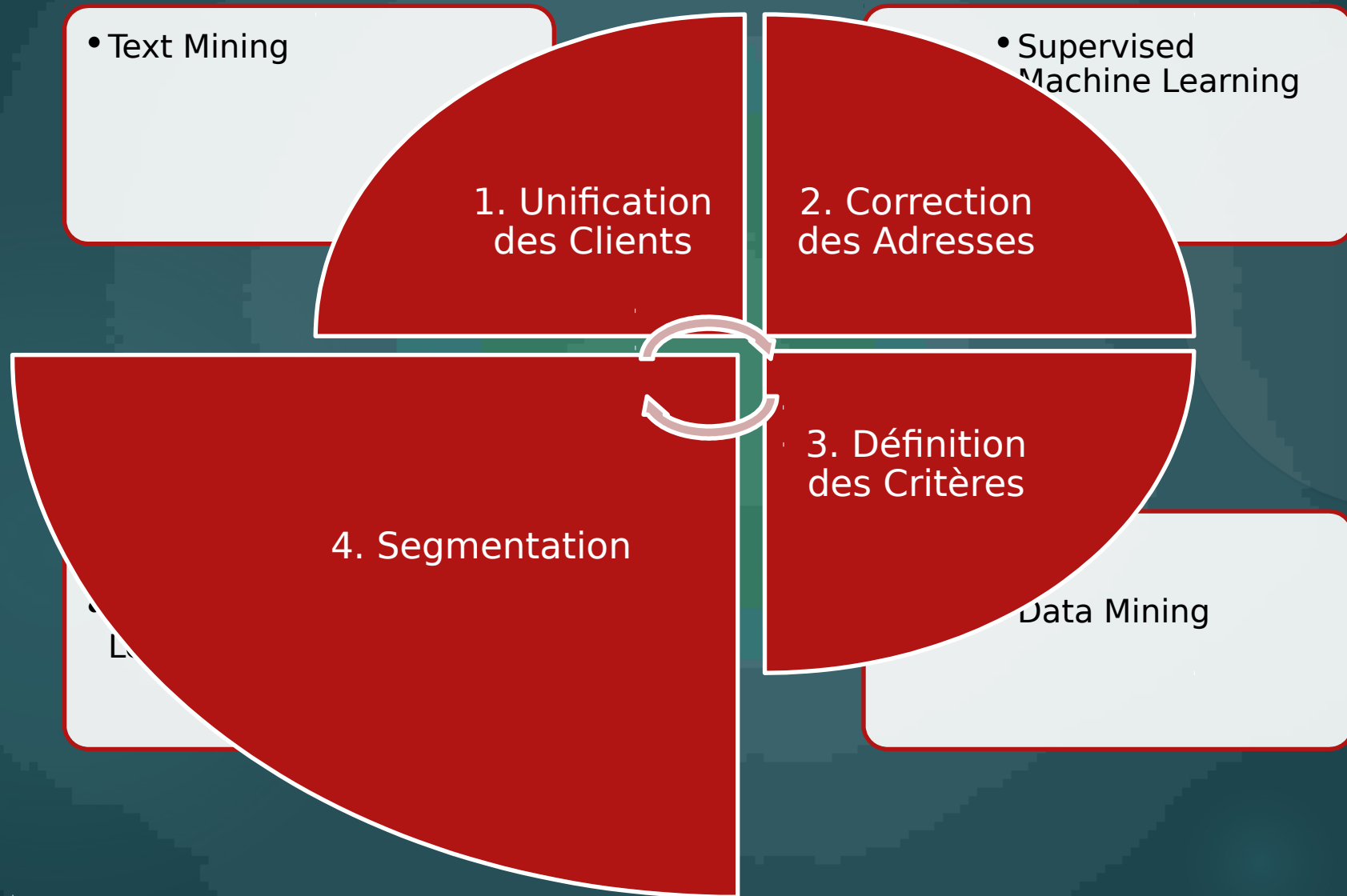
Extrait des données avec les critères définis...

TIC_ENT_ID	Coord	Civ	Région	Magasin	NewsLet	Age	DiversitePro	Porteur	Tenure	Récence	MontWeb	MontPub	MontMeubl	MontB
10003875	ZAEPFEL DANIEL	Mr	Rhône Alpes	138	0	23	4	1	84	1419	0	0	94.3	5.7
10040475	ZAIDI NADIA	Mme	Centre	819	0	20	9	1	637	887	0	0	93.7	14.66
10045167	ZAJA MR	Mr	Rhône Alpes	848	0	NA	1	0	0	1496	0	0	100	
10046717	ZAIDOUNI MALIKA	Mme	Aquitaine	848	0	29	6	1	842	998	0	0	78.66	30.56
10063137	ZAIRE DAVID	Couple	Rhône Alpes	848	0	22	10	1	387	1452	0	0	102.76	0.07
10072003	ZAFFINI	Couple	Languedoc Rou	848	0	NA	11	0	1087	1006	0	0	96.88	3.73
10074339	ZABOURI	Mme	Rhône Alpes	848	0	NA	4	0	0	1617	0	0	83.39	16.61
10074774	ZAITI PATRICIA	Mme	Languedoc Rou	819	0	NA	2	0	0	1552	0	0	73.57	26.43
10106508	ZAHOUSSAMA	Mme	Languedoc Rou	848	0	27	6	1	1341	889	0	0	93.5	6.5
10110640	ZAHABDELMAJID	Mr	Bretagne	848	0	28	3	1	59	1930	0	0	100	
10112900	ZAITER MME	Mme	Bretagne	848	0	NA	2	0	0	1839	0	0	98.56	1.44
10113641	ZAGHOUBANI	Couple	Rhône Alpes	848	0	NA	4	0	358	1500	0	0	96.77	3.23
10128055	ZAFFINI CHANTAL	Mme	Centre	848	0	NA	1	0	0	2188	0	0	100	
10158051	ZAHLAOUI JULES	Mr	Franche Comté	819	0	NA	1	0	0	1649	0	0	100	
10168042	ZADRI MME	Mme	Aquitaine	848	0	NA	1	0	0	1874	0	0	100	

Data Science Project Lifecycle



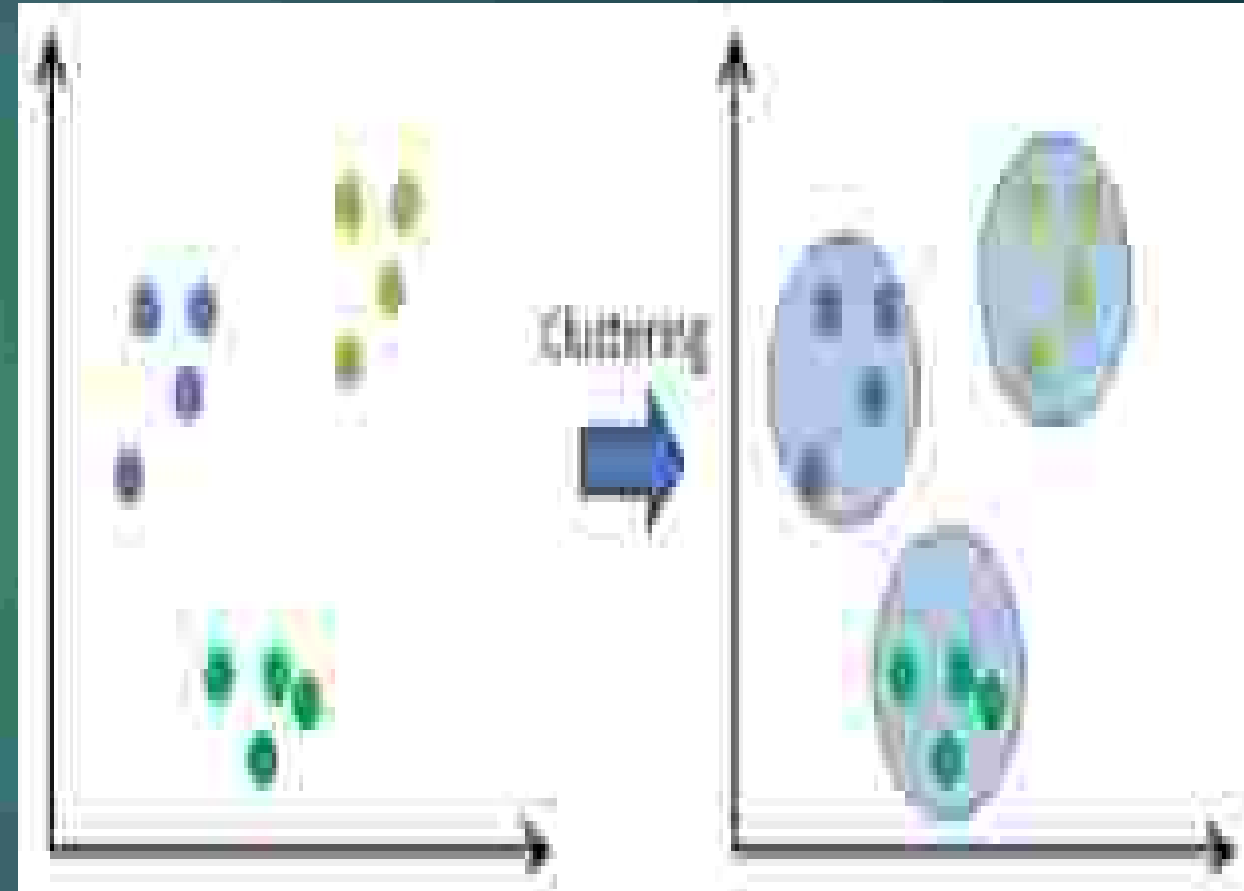
AXES DU PROJETS



AXES DU PROJET

CLUSTERING AUTOMATISÉ

- Le Clustering est un processus de regroupement à base de critères
- Méthode se basant sur une mesure de similarité entre les objets à regrouper
- Nombre de clusters déterminé par l'analyste (choix non arbitraire !)
- Construire une collection d'objets :
 - Similaires au sein d'un même groupe
 - Dissimilaires quand ils appartiennent à des groupes différents



AXES DU PROJET

CLUSTERING AUTOMATISÉ

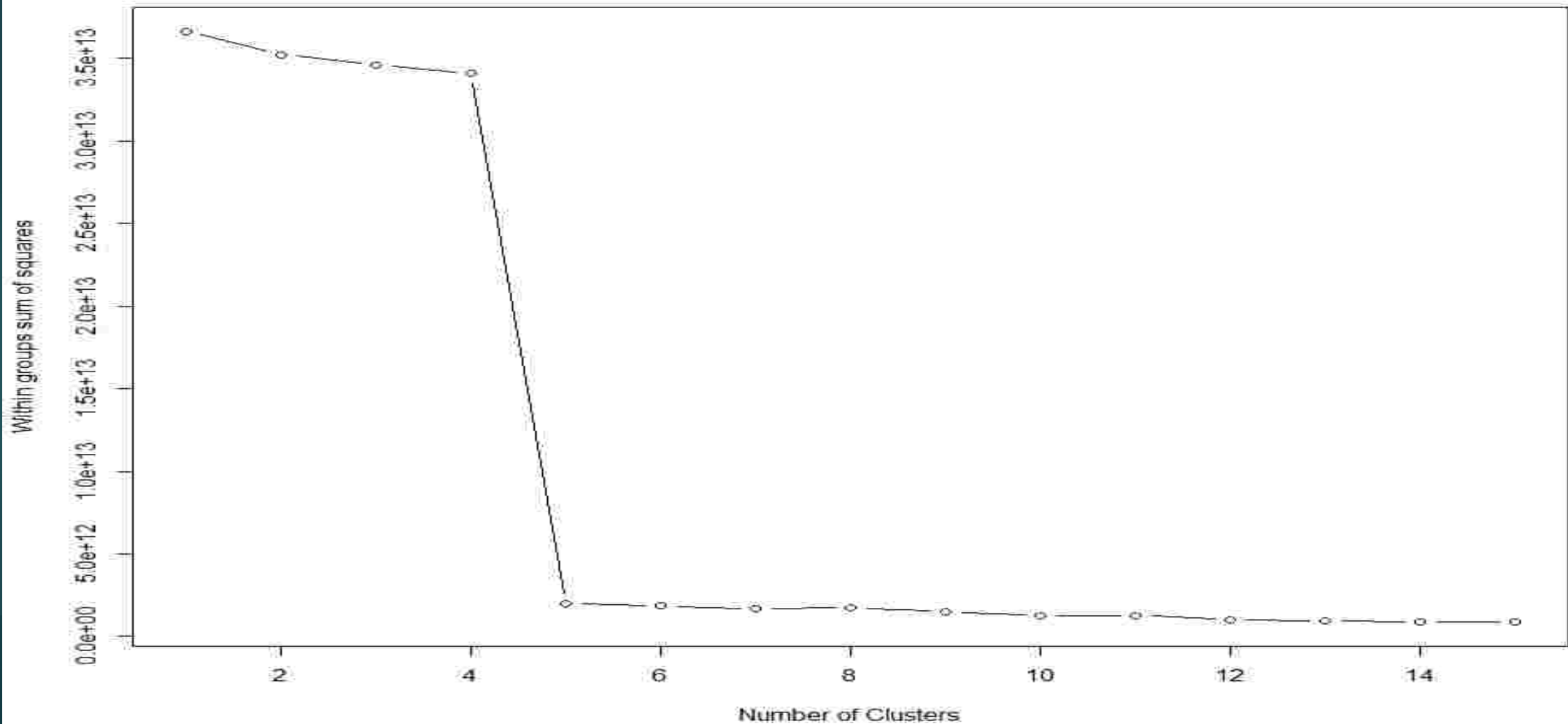


Critères de Clustering	
NewsLet	Age
DiversiteProduits	Porteur
Tenure	Recence
MontWeb	MontPub
MontPub	MontBoutique
MontTopSelling	MontPresta
CarteFid	CarteFidPlus
TotalMont	NbWeb
NbPub	NbMeuble
NbBoutique	NbTopSelling
NbTotal	NbPassages
NbRetours	HorsWknd
PaiPersonnel	PaiCredit
PaiBons	Remise
F022	F023
TotalPaiement	MargeMeuble
MargeBoutique	

AXES DU PROJET

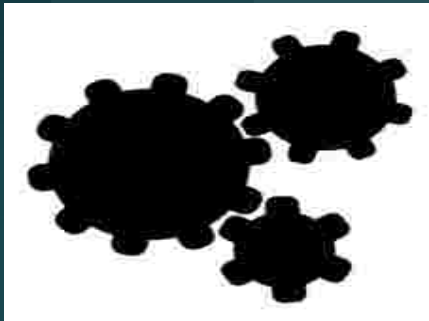
CLUSTERING AUTOMATISÉ

Combien de Groupes ??



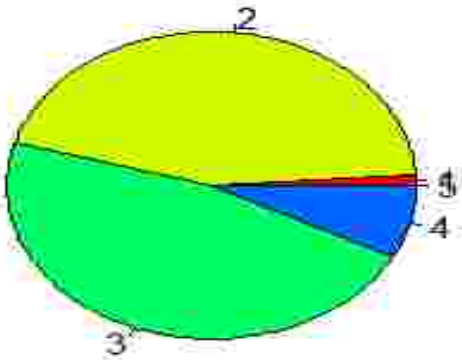
AXES DU PROJET

CLUSTERING AUTOMATISÉ



Numéro du Groupe	1	2	3	4
Nombre de Clients	26797	1144268	1206319	199831
Pourcentage	1,04	44,40	46,81	7,75

Distribution des clients par Groupe



AXES DU PROJET

CLUSTERING AUTOMATISÉ

Examinons les
caractéristiques de
chaque cluster de
clients...



AXES DU PROJET

CLUSTERING AUTOMATISÉ

GROUPE 1 (1,04%) :

- Large diversité de produits
- La moitié des clients sont des porteurs de carte
- Longue tenure
- La moitié ont disparu depuis plus de deux ans
- Achat élevé d'articles boutique
- Quantité d'articles achetés très élevée (75% des clients ont acheté plus de 20 articles)
- Nombre de passages élevé (75% des clients sont passés au moins 4 fois)
- Nombre de retours très faible (plus de 75% des clients n'ont pas fait de retours)
- Achats étalés sur la semaine
- Utilisation très faible de F023 et de carte de crédit, encore plus faible de bons, encore plus faible de F022

Clients fidèles et potentiels perdus !

AXES DU PROJET

CLUSTERING AUTOMATISÉ

GROUPE 2 (44,40%) :

- Faible Diversité de produits (75% des clients ont acheté au mieux 4 sous-familles)
- Non porteurs de cartes (uniquement 18.74% sont porteurs)
- La moitié sont venus une seule fois
- Plus de 75% ne sont venus ça fait plus d'un an
- Plus de 25% n'achètent que du meuble
- Faible quantité achetée (la moitié ont pris au mieux 3 articles)
- 75% des clients sont passés une seule fois
- 25% ont fait leur achat hors week-end
- Uniquement 4.31% des clients ont fait des retours
- Presque la moitié font un financement purement personnel
- Pas de F022/F023

Clients non fidèles et non potentiels mais possible de les activer

AXES DU PROJET

CLUSTERING AUTOMATISÉ

GROUPE 3 (46,81%) :

- Très faible Diversité de produits (75% des clients ont acheté au mieux 3 sous-familles)
- Non porteurs de cartes (uniquement 9.73% sont porteurs)
- Tous les clients ne sont venus plus de 3 ans
- Très faible tenure (75% des clients ont au mieux une tenure de 3 jours)
- La moitié n'achètent que du meuble
- Faible quantité achetée (75% des clients ont acheté au mieux 5 articles)
- La moitié sont venus une seule fois
- Uniquement 2.75% ont fait des retours
- 25% ont fait leur achat pendant le weekend, 25% ont fait leur achat hors le weekend
- 25% des clients font un financement purement personnel
- Pas d'achat Web

Clients non fidèles et non potentiels perdus

AXES DU PROJET

CLUSTERING AUTOMATISÉ

GROUPE 4 (7,75%) :

- Diversité de produits élevée (75% des clients achètent au moins 5 sous-famille)
- La moitié sont des porteurs de cartes
- Large tenure (La moitié ont une tenure de presque 2 ans)
- Clients récents (la moitié des clients ont une récence de moins de 2 ans)
- Achat d'articles en Pub
- Achat top selling
- Achat équilibré de meuble et boutique
- Quantité achetée élevée (la moitié achète plus de 15 articles)
- Nombre de passages élevés (la moitié sont passés plus de 4 fois)
- Achat étalé sur la semaine
- Financement varié (utilisation du crédit, bons, F022 et F023)

Clients fidèles et actifs et récents à fidéliser davantage