



# L'IA au service des agents immobiliers

- Silicon Valley

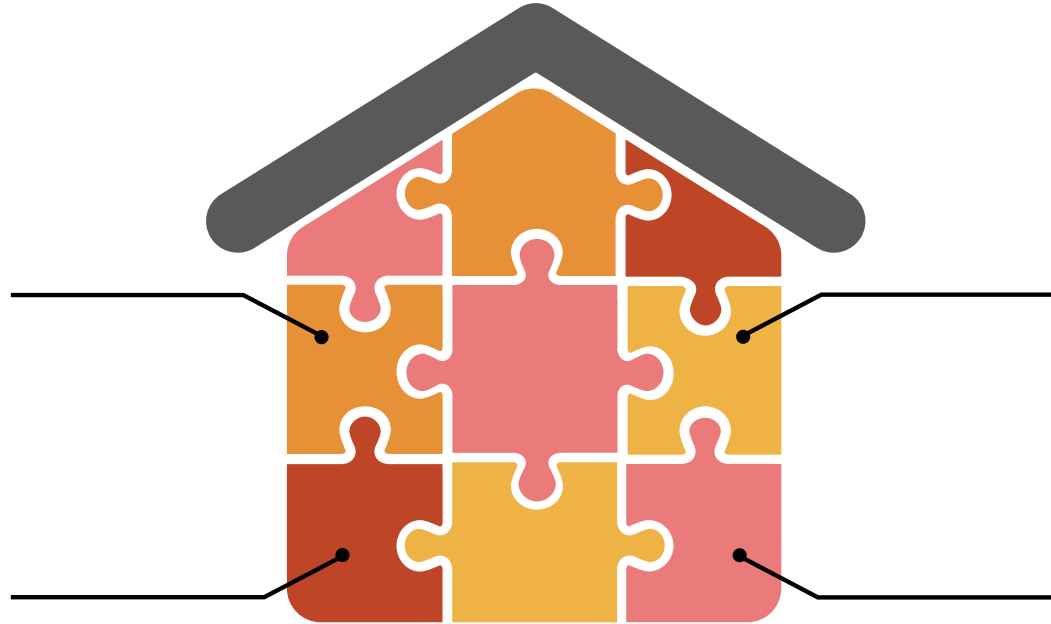
# Les données

## La localization :

Longitude,  
Latitude, Ocean  
proximity

## Les habitants :

Population, Median  
income

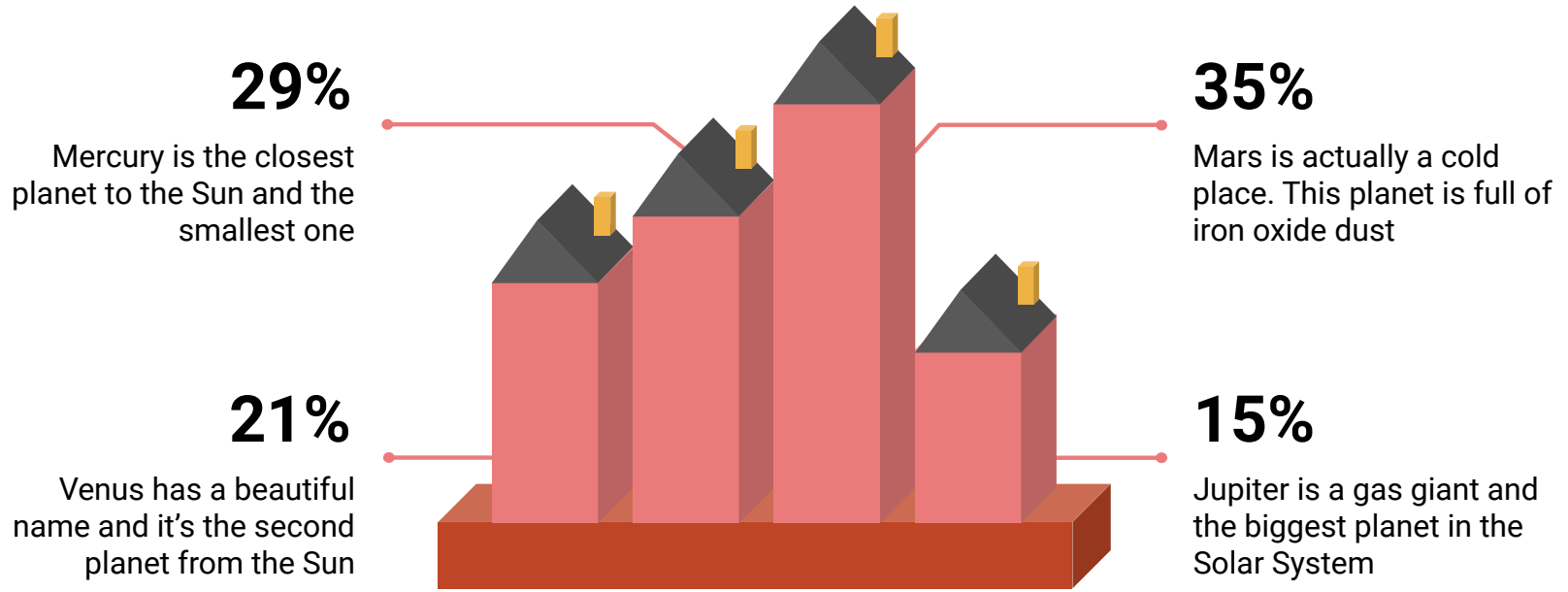


Households

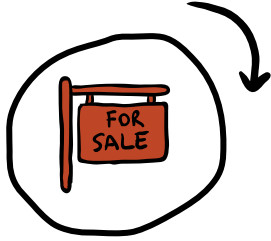
## Les habitations:

Housing median  
age, Total rooms,  
Total bedrooms,  
Median house  
value

# Real Estate Infographics

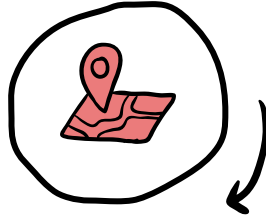


# NOTEBOOK



## EDA

Exploration / Missing  
Values / Encoding /  
Visualisation



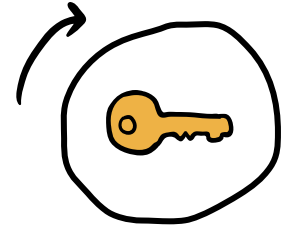
## INFERENCE

Analyse d'inférence  
statistique avec  
Statsmodels



## MODEL

Présentations de toutes  
les itérations



## PREDICT

Pré-cleaning et utilisation  
du model

# EDA



## Encoding

Encodage de la colonne  
ocean\_proximity avec  
get\_dummies(librairie Pandas)



## Dropping

Suppression de la colonne  
Unnamed:0 qui n'a aucune utilité



## Missing Values

Traitement des 176 données  
manquantes sur total\_bedrooms  
avec le KNN Imputer afin de  
remplacer les valeurs manquantes  
par la moyenne de ses voisins

# INFERENCE

Test de Pearson



**CORRELATION**

Hypothèse gaussienne



**KS TEST**

R-squared = 64,7%



**OLS**

Variable expliquée



**PCA**

Elimination de features



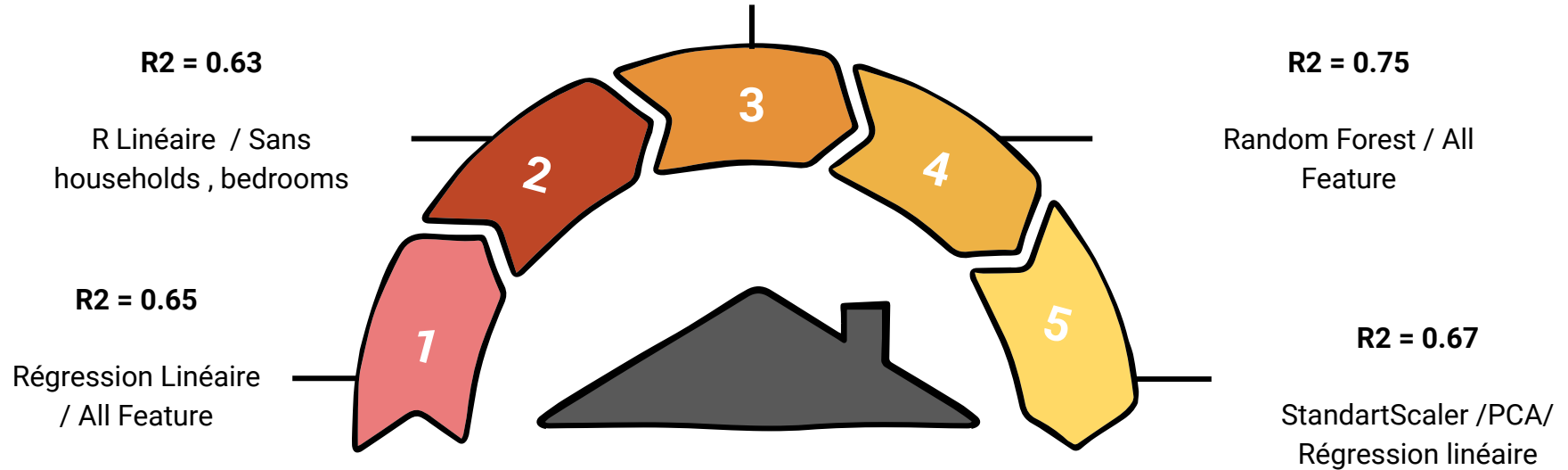
**CERCLE DE  
CORRÉLATION**

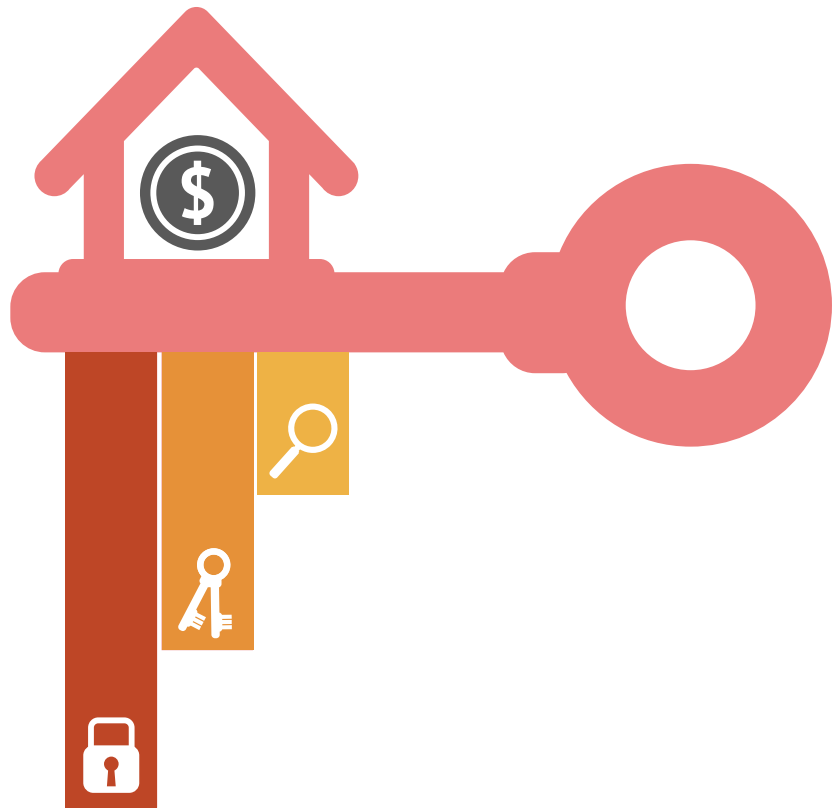


# MODEL

**R2 = 0.62**

Régression Linéaire / Sans Outliers





# PREDICT

KNN Regressor a été l'algorithme le plus adéquat en utilisant les points voisins suggérant le meilleur prix de vente

1

$R^2 = 0.82$

2

Mean Absolute Error = 32 876 \$

3

Mean Absolute Percentage Error = 17,8%



# AXE D'AMELIORATION

Meilleur gestion des outliers



Explorer et enrichir mon dataset avec des features pertinentes



Refactorisation du code



Poussé encore plus l'inférence



Faire plus d'itérations

