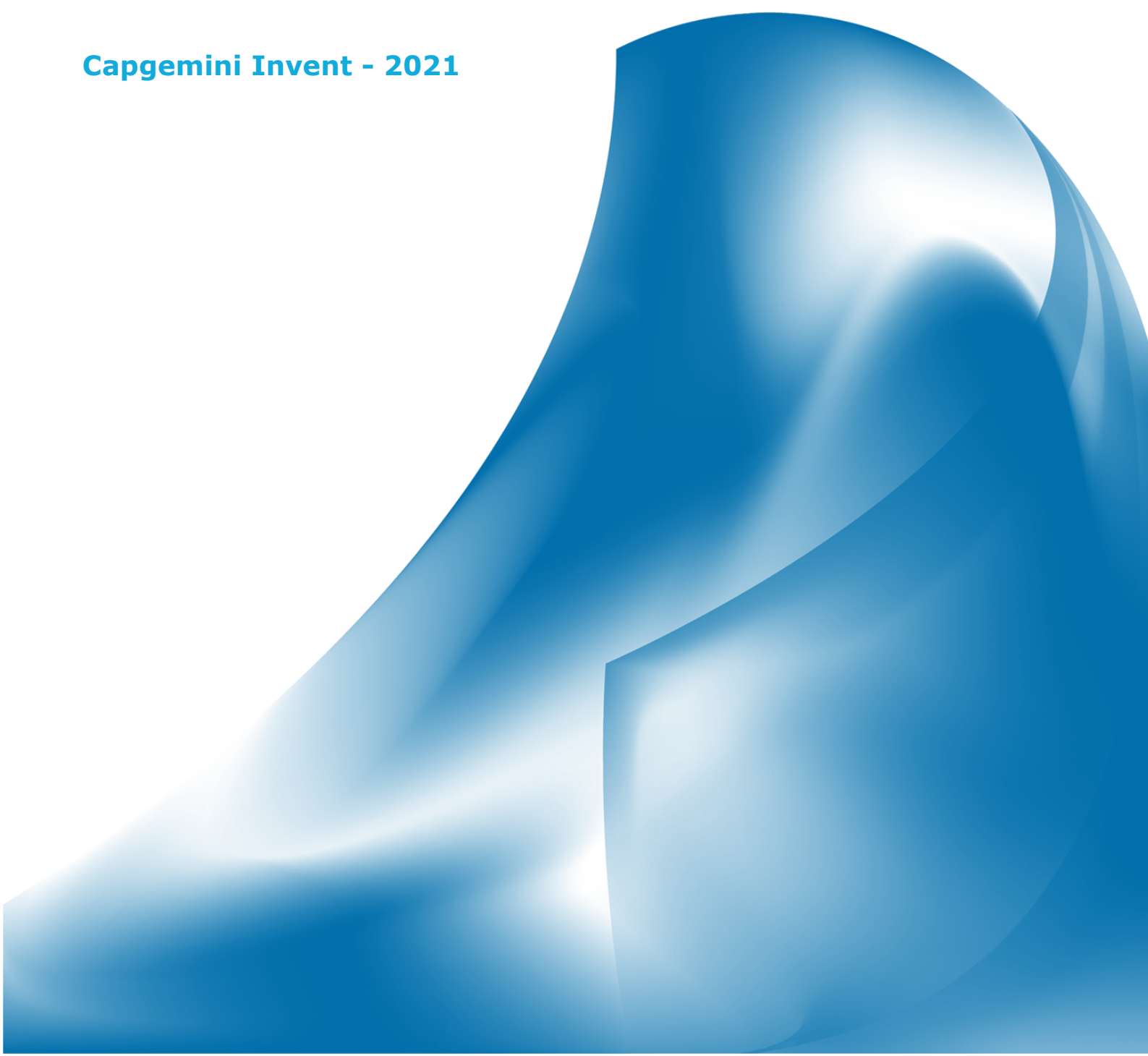


RECRUITMENT TEST MACHINE LEARNING

Capgemini Invent - 2021





INSTRUCTIONS

Présentation

Ce test a pour but de vous mettre en situation d'un consultant data scientist sur mission afin d'évaluer vos compétences et de mieux vous connaître. Cette version du test a été choisie en fonction de votre expérience, de votre séniorité et de vos réponses au questionnaire d'auto-évaluation.

Le délai qui vous est laissé pour effectuer ce test est d'une semaine. Le temps de travail recommandé est entre 2 heures et 20 heures, en fonction de vos expertises et de vos contraintes personnelles et professionnelles. Il vous sera demandé de nous communiquer un temps de travail passé sur le projet pendant la restitution, de façon déclarative, afin d'adapter le niveau d'attente.

Livrables

Pour défendre votre proposition de solution, les livrables attendus sont :

- Un dossier contenant du code fonctionnel et un readme permettant de l'exécuter, à envoyer la veille de la restitution. Le code sera lu et analysé.
- Un support de présentation de format libre (markdown, jupyter notebook, webapp, powerpoint, ...) permettant de présenter votre proposition à un public mixte technique et non technique.
- Une présentation à l'oral en séance

Critères d'évaluation

Les points suivants seront évalués :

- Avant la restitution
 - Approche technique
 - Qualité du code
- Pendant la restitution
 - Pertinence de l'approche de la demande du client
 - Méthodologie machine learning
 - Capacité à présenter et posture
 - Réponses aux questions business et techniques
- Après la restitution
 - Discussions plus générales sur l'approche et la méthodologie
 - Discussions et questions en machine learning, sur ou en dehors de l'étude de cas considérée

Le niveau d'attente est adapté en fonction de votre expérience et séniorité.

Avertissements

L'énoncé de ce test de recrutement est la propriété intellectuelle de Capgemini et le candidat ne doit pas le partager.

Le travail produit par le candidat lors de ce test de recrutement est la propriété intellectuelle du candidat et il ne sera pas utilisé au sein de Capgemini à d'autre fins que le test de recrutement du candidat.



ETUDE DE CAS 1

Situation

Le client est le responsable des outils et méthodes d'un réseau d'agences immobilières historique en Haute Garonne (31) qui vend principalement des maisons et des appartements d'occasion. Le gouvernement vient de mettre en open data des données relatives aux données de valeurs foncières. Le client souhaite tirer parti de ces données pour fournir un outil à ses agents afin d'évaluer au plus juste la valeur de vente d'un bien immobilier en fonction des caractéristiques connues du bien. Leur outil actuel est basé sur des règles métiers et l'expérience professionnelle de chacun des agents. Le client est convaincu qu'une nouvelle approche peut apporter de la valeur à son entreprise et lui permettre de rester en concurrence face aux néo agences digitales. Cependant le directeur et fondateur de l'entreprise a plus confiance dans l'expérience professionnelle et individuelle de ses agents et demande à être convaincu avant d'engager plus de frais sur cette idée. Le client vous demande alors d'évaluer cette possibilité en travaillant sur une preuve de concept et en préparant un argumentaire démontrant la valeur, ou l'absence de valeur, de cette solution, et quelles seraient les prochaines actions à envisager.

Jusqu'alors, les agents utilisent leur outil basé sur des règles métiers ainsi que leur expérience pour estimer un prix de vente net acheteur. Il est préférable pour un agent immobilier de prévoir un prix plus haut que le marché plutôt qu'un prix plus bas.

Data

Les données disponibles sont décrites ci-dessous. Il n'est pas forcément nécessaire d'utiliser toutes les données fournies. Il est également autorisé, mais pas nécessaire, d'utiliser des données externes.

- **agence.csv** : La liste des biens vendus par le réseau d'agences du client avec des données complémentaires
 - **date_mutation** : date de vente telle qu'enregistrée dans le cadastre
 - **valeur_fonciere** : Montant de vente effective du bien
 - **code_postal** : Code postal
 - **nom_commune** : Commune
 - **id_parcelle** : Identifiant de la parcelle cadastrale
 - **type_local** : Appartement ou Maison
 - **surface_reelle_bati** : Surface au sol du bati
 - **nombre_pieces_principales** : Nombre de pieces principales du logement
 - **agent** : numéro identifiant de l'agent responsable de la vente
 - **evaluation_agent** : Évaluation de l'état du bien par l'agent
 - **prix_cible** : Prix de vente final estimé par l'agent chargé du bien par la méthode actuelle
 - **prix_vente_initial** : Le prix de vente de mise sur le marché par l'agent, basé sur l'estimation du prix de vente cible avec une marge de négociation
- **insee_commune.csv** : Données INSEE prétraitées à l'échelle de la commune, issue de la source suivante : <https://www.insee.fr/fr/statistiques/2521169#consulter>
- **localisation_commune.csv** : Données de localisation des communes issues de la base officielle des codes postaux
 - https://datanova.legroupe.laposte.fr/explore/dataset/laposte_hexasmal/information/?disjunctive.code_commune_insee&disjunctive.nom_de_la_commune&disjunctive.code_postal&disjunctive.libell_d_acheminement&disjunctive.ligne_5



- france.csv : Les données de valeur foncières de 2015 à 2020 sur toute la France, issues des bases cadastrales open data, enrichies des données INSEE à l'échelle du carreau naturel.
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2015/full.csv.gz>
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2016/full.csv.gz>
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2017/full.csv.gz>
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2018/full.csv.gz>
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2019/full.csv.gz>
 - <https://cadastre.data.gouv.fr/data/etalab-dvf/latest/csv/2020/full.csv.gz>
 - Reference des données cadastrales :
 - <https://static.data.gouv.fr/resources/demandes-de-valeurs-foncieres/20191220-102114/notice-descriptive-du-fichier-dvf.pdf>
 - Données INSEE : <https://www.insee.fr/fr/statistiques/4176281?sommaire=4176305>
- Idées de données externes complémentaires
 - Coordonnées GPS du centre-ville de Toulouse : Lat 43.603, Lon 1.443
 - Densité de population dans les communes : <https://www.insee.fr/fr/information/2114627>

ABOUT CAPGEMINI

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 200,000 team members in over 40 countries. The Group reported 2016 global revenues of EUR 12.5 billion.

Learn more about us at www.capgemini.com



People matter, results count.