

◆ 2022-1 겨울방학

Sooktat

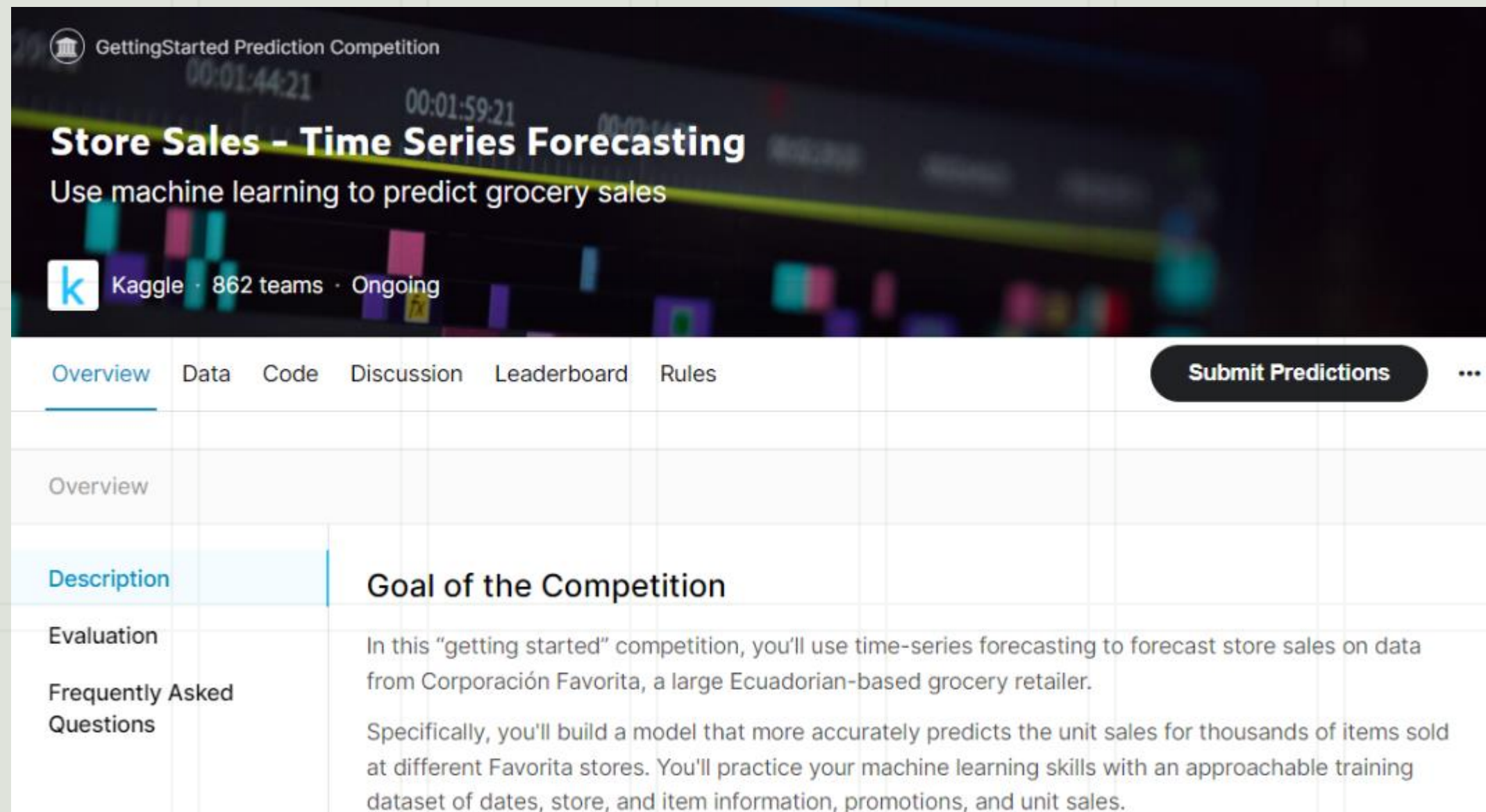
kaggle data analysis project

2조 중간발표

고나경, 김윤진, 김은호, 정소영, 최윤서

Competition

시계열 상품 구매 데이터를 이용한 상품 수요 예측



GettingStarted Prediction Competition

Store Sales - Time Series Forecasting

Use machine learning to predict grocery sales

Kaggle · 862 teams · Ongoing

Overview Data Code Discussion Leaderboard Rules

Submit Predictions

Overview

Description

Goal of the Competition

In this "getting started" competition, you'll use time-series forecasting to forecast store sales on data from Corporación Favorita, a large Ecuadorian-based grocery retailer.

Specifically, you'll build a model that more accurately predicts the unit sales for thousands of items sold at different Favorita stores. You'll practice your machine learning skills with an approachable training dataset of dates, store, and item information, promotions, and unit sales.

<https://www.kaggle.com/c/store-sales-time-series-forecasting>

data

	date	type	locale	locale_name	description	transferred
0	2012-03-02	Holiday	Local	Manta	Fundacion de Manta	False
1	2012-04-01	Holiday	Regional	Cotopaxi	Provincializacion de Cotopaxi	False
2	2012-04-12	Holiday	Local	Cuenca	Fundacion de Cuenca	False
3	2012-04-14	Holiday	Local	Libertad	Cantonizacion de Libertad	False
4	2012-04-21	Holiday	Local	Riobamba	Cantonizacion de Riobamba	False

holidays_events.csv

	store_nbr	city	state	type	cluster
0	1	Quito	Pichincha	D	13
1	2	Quito	Pichincha	D	13
2	3	Quito	Pichincha	D	8
3	4	Quito	Pichincha	D	9
4	5	Santo Domingo	Santo Domingo de los Tsachilas	D	4

stores.csv

	date	dcoilwtico
0	2013-01-01	NaN
1	2013-01-02	93.14
2	2013-01-03	92.97
3	2013-01-04	93.12
4	2013-01-07	93.20

oil.csv

	date	store_nbr	transactions
0	2013-01-01	25	770
1	2013-01-02	1	2111
2	2013-01-02	2	2358
3	2013-01-02	3	3487
4	2013-01-02	4	1922

transaction.csv

	id	date	store_nbr	family	sales	onpromotion
0	0	2013-01-01	1	AUTOMOTIVE	0.0	0
1	1	2013-01-01	1	BABY CARE	0.0	0
2	2	2013-01-01	1	BEAUTY	0.0	0
3	3	2013-01-01	1	BEVERAGES	0.0	0
4	4	2013-01-01	1	BOOKS	0.0	0

train.csv

	id	date	store_nbr	family	onpromotion
0	3000888	2017-08-16	1	AUTOMOTIVE	0
1	3000889	2017-08-16	1	BABY CARE	0
2	3000890	2017-08-16	1	BEAUTY	2
3	3000891	2017-08-16	1	BEVERAGES	20
4	3000892	2017-08-16	1	BOOKS	0

test.csv

진행 계획

1월


SUN	MON	TUE	WED	THU	FRI	SAT
						15 첫 회의
16	17	18	19	20	21	22
데이터 탐색 및 EDA						
23	24 중간발표	25	26	27	28	29
EDA 및 feature engineering						
30	31					

2월



SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
feature engineering 및 모델 구성						
6	7	8	9	10	11	12
모델 구성 및 모델 튜닝						
13	14	15	16	17	18	19
모델 구성 및 모델 튜닝						
20	21	22	23	24	25	26
프로젝트 마무리 및 ppt 제작						
27	28 최종 발표					

진행 상황

github 개설 및 데이터 정보 확인

 YunSeo00 Update readme.md d351ae6 yesterday 27 commits

ProjectCode/20220122	Update readme.md	yesterday
README.md	Update README.md	yesterday

 README.md 

SookTat-Project_team2-2022-1

- 숙택의 2022-1학기 kaggle competition project 진행 과정을 기록한 저장소입니다.

Kaggle Project

- Competitions: [Store Sales - Time Series Forecasting](#)
- Member : 고나경, 김윤진, 김은호, 정소영, 최윤서

20220122

- [고나경](#)
-
-
- 김은호
 - [프로그램 코드](#)
 - [데이터 설명 파일](#)
- 최윤서

개요

제품 수요를 예측, Corporación Favorita에서 판매되는 제품군에 대한 매출을 예측

교육 데이터 : 날짜, 상점 및 제품 정보, 해당 품목의 프로모션 여부, 판매 번호

- 목적 : 제품 수요 예측을 통해 소매업체의 제품 공급 도움
- 예상 효과 : 재고 과잉 관련 음식물 쓰레기 감소, 고객 만족도 상승, 지역 상점의 필요성 확인

데이터 확인

`train.csv` ; 학습 데이터 .

- id
- date 판매된 날짜
- store_nbr 제품이 판매되는 상점
- Family 판매되는 제품 유형
- sales 지정된 날짜에 특정 상점에서 제품군에 대한 총 판매
- onpromotion 은 특정 날짜에 매장에서 판촉된 제품군의 총 항목 수

`test.csv` ; 검증 데이터

- id, date, store_nbr, Family, onpromotion
- sales (X) 예측

진행 상황 EDA

