
주식거래 데이터 속 이상치

박현우

이상치

- **Point Anomaly** : 정상데이터의 분포로부터 완전히 벗어난 데이터 (가장 많은 연구진행됨)
 - **Contextual Anomaly** : 데이터 분포 자체에는 이상점이 없으나 데이터 흐름 또는 맥락상 정상이 아닌 데이터 (특정 시점에는 특정데이터가 나와야하는데 다른데이터 나오는경우)
 - **Group Anomaly** : 단일데이터만 봤을땐 정상처럼 보이지만 하지만 두데이터를 모아서 보았을때 비정상으로 판단되는 이상치 종류 (몇일에 걸쳐 계속 같은 결제반복, 그래프가 미미하지만 지속적으로 우상향하는 진동 센서 -> 이상치)
-

이상치의 예시

➤ 포스코 인터내셔널의 일별 수익률 최대인 날짜, 거래량이 최대인 날짜의 기사

```
data[data['pct_change'] == data['pct_change'].max()]
```

	Date	Open	High	Low	Close	Adj Close	Volume	시장	diff	pct_change
Date										
2021-02-04	2021-02-04	19350.0	24200.0	18850.0	22600.0	21807.017578	18348535	코스피	0.236726	0.170984

```
data[data['Volume'] == data['Volume'].max()]
```

	Date	Open	High	Low	Close	Adj Close	Volume	시장	diff	pct_change
Date										
2021-02-04	2021-02-04	19350.0	24200.0	18850.0	22600.0	21807.017578	18348535	코스피	0.236726	0.170984

[포스코인터내셔널, 전기차 부품 사업 기대감에 17% 급등..기관 130억 샀다](#)

특징주

포스코인터내셔널이 전기차 부품 사업 기대감에 급등했다. 4일 포스코인터내셔널은 전일 대비 3300원(17.10%) 오른 2만2600원에 마감했다. 장중 주가는 2만4200원까지 올라 52주 최고가를 경신했다. 이날 기관은 홀로 포스코인터내셔널 130억원을 사들이며 주머니투데이 | 2021.02.04 | 다음뉴스

이상치의 예시

➤ 포스코인터내셔널의 5일 이동평균(요일 편향 반영)과 일별 수익률의 최대 차이(양수) 를 가진 날짜의 기사

	Date	Open	High	Low	Close	Adj Close	Volume	시장	diff	pct_change	5_ma	mov_diff
Date												
2021-07-23	2021-07-23	23650.0	27300.0	23650.0	27200.0	26245.615234	9447199	코스피	0.134191	0.150106	0.033136	0.11697

[포스코인터내셔널, 사상 최대 분기 실적에 10%대 강세](#)

특징주

포스코인터내셔널이 사상 최대 분기 실적에 장중 10%대 급등 중이다. 하반기 미얀마 가스전 이익 회복이 기대된다는 증권사 전망도 긍정적으로 작용한 것으로 풀이된다. 23일 오후 3시8분 포스코인터내셔널은 전일 대비 3200원(13.53%) 오른 2만6850원에 거래되고...
머니투데이 | 2021.07.23 | 다음뉴스

[포스코인터내셔널, 2분기 매출 8조5245억 '사상 최대'](#)

포스코인터내셔널이 올 2분기 사상 최대 분기매출 실적을 올렸다. 포스코인터내셔널은 21일 2분기 잠정실적 공시를 통해 사상 최대 분기 매출인 8조5245억원, 영업이익 1700억, 순이익 1262억을 달성했다고 밝혔다. 2분기 매출액과 영업이익은 전년 동기 대비 각각 6...
경북일보 | 2021.07.22

| [포스코인터내셔널, 2분기 매출 8조5245억원...분기 ...](#) 이뉴스투데이 | 2021.07.22



[\(포스코 IR\) 자회사 경영실적, 포스코인터내셔널 '철강 판매량 증가'](#)

중심 트레이딩 호조로 매출액 및 영업이익 증가했다. 이는 철강 판매량 증가했고 단가 상승이 이뤄졌기 때문이다. 2분기 포스코인터내셔널의 매출액은 7조7,470억원, 영업이익은 1,090억원으로 각각 직전 분기 대비 1조4,440...
철강금속신문 | 2021.07.22

| [\(포스코 IR\) 주요 해외 자회사 실적 '대부분 수익...](#) 철강금속신문 | 2021.07.22

➤ 포스코인터내셔널의 5일 이동평균과 일별 수익률의 최대 차이(음수) 를 가진 날짜의 기사

	Date	Open	High	Low	Close	Adj Close	Volume	시장	diff	pct_change	5_ma	mov_diff
Date												
2022-03-08	2022-03-08	23000.0	23050.0	20800.0	20800.0	20800.0	2698804	코스피	0.108173	-0.0837	0.006657	-0.090357



[포스코인터내셔널, 우크라이나 난민 위해 50만달러 기부](#)

포스코인터내셔널(대표이사 주시보)이 우크라이나 난민을 위해 기부에 동참했다. 회사는 앞서 우크라이나 현지 진출 이후 지역사회와의 상생활동을 지속해오고 있다. 포스코인터내셔널은 대한적십자사(대표이사 회장 신희영)를...
더벨류뉴스 | 2022.03.08

프로젝트를 위한 이상치 정의

- 시장 수익률과 비교했을 때 지속적인 우상향을 하기 시작한 시점 근처
- 시장 거래대금 흐름과 종목 거래대금이 극단적으로 다른 point 들
- 개장 직후 상한가, 하한가 종목 (전날이나 주말에 기업 소식이 원인일 가능성 큼)
 - 거래량이 적을 확률이 크다 -> 일별 수익률이 Input Data로 필요한 이유

기존 코스피 거래량의 문제점

- 코스피 거래량 데이터는 코스피 전종목의 거래량 데이터를 가중치를 부여하지 않고 단순 합을 이용해 계산된 것입니다.
- 가중치 부여 방안 : S = 코스피 전종목 일별 종가의 합 , V_i = i 종목 일일 종목거래량
 - W_i = 일일 종목 i 가중치 = 종목 i 종가 / S
 - 일일 코스피 거래량 = $\text{sum}(W_i * V_i)$

거래대금과 거래량

➤ 거래량 vs 거래대금

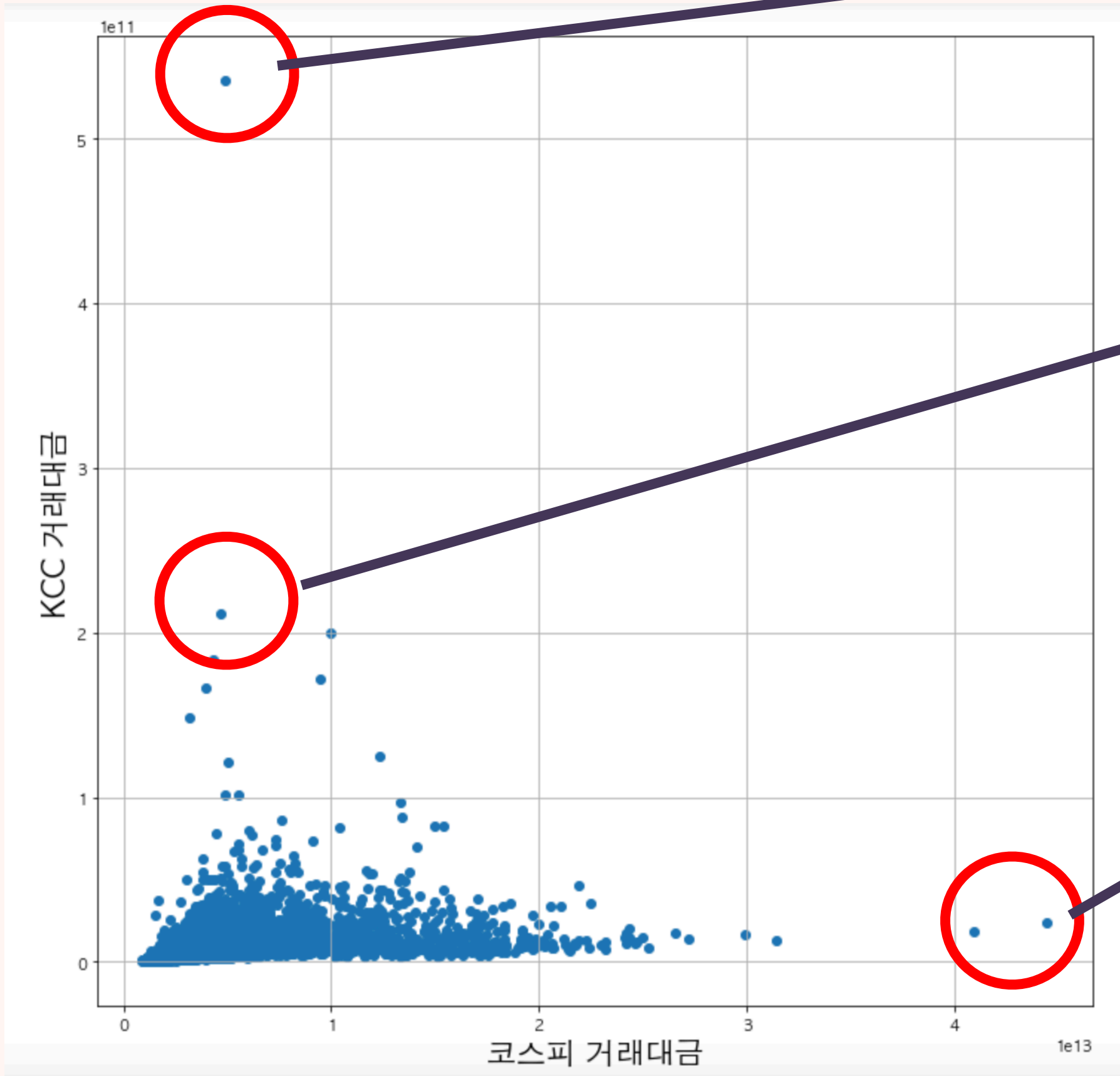
Random Sampling을 통해 코스피 200으로부터 30개 종목 sampling을 30번 반복했을 때

코스피 거래대금과 종목 거래대금 사이 샘플 상관계수 평균들의 상관계수 평균 0.6565686121636848
코스피 거래량과 종목 거래량 사이 샘플 상관계수 평균들의 상관계수 평균 0.5892800353800483

Random Sampling을 통해 코스피 200으로부터 30개 종목 sampling을 100번 반복했을 때

코스피 거래대금과 종목 거래대금 사이 샘플 상관계수 평균들의 상관계수 평균 0.6545953882296937
코스피 거래량과 종목 거래량 사이 샘플 상관계수 평균들의 상관계수 평균 0.587893990278979

	날짜	시가	고가	저가	종가	거래량	거래대금	등락률	diff	pct_change
날짜										
2014-11-20	2014-11-20	527000	529000	510000	511000	1035233	535393669000	-6.07	0.037182	-0.060662
2016-06-15	2016-06-15	365500	393500	364500	378000	581211	211506385250	-0.26	0.076720	-0.002688



KCC, 3천억 규모 현대중공업 주식 매입 결정..지원 나서나

- 정몽진 KCC회장-정몽준 현대중공업 대주주 사촌지간 [헤럴드경제=박수진 기자] 범현대가인 KCC가 부동산을 3000억원 규모의 현대중공업 주식을 매입하기로 했다. 정몽진 KCC회장과 정몽준 현대중공업 대주주는 사촌 지간이다. 21일 금융감독원 전자공시시스템…
헤럴드경제 | 2014.11.21 | 다음뉴스

- KCC, 현대重 지분 매입..사촌형 정몽준 지원 나서나 한국경제 | 2014.11.21 | 다음뉴스
 - KCC, 현대重 지원 나서나…주식 매입에 3천억원 투입 EBN | 2014.11.21
 - KCC, 3천억 규모 토지·건물 처분 연합뉴스 | 2014.11.20 | 다음뉴스
- 관련뉴스 6건 전체보기 >

현대미포조선, 1400억원대 KCC 지분 전량 처분키로

이날 장 마감 뒤 블록딜 추진.. 총 매각금액 1413억~1472억원 예상
기자] [이날 장 마감 뒤 블록딜 추진… 총 매각금액 1413억~1472억원 예상] 현대중공업그룹 계열사인 현대미포조선이 보유중인 KCC 지분을 모두 처분한다. 현대중공업그룹의 경영개선계획에 따라 앞으로 계열사 보유지분에 대한 매도가 지속될 것으로 전망된다. 14일…
머니투데이 | 2016.06.14 | 다음뉴스

- 현대미포조선, KCC 지분 전량 처분 서울경제 | 2016.06.14 | 다음뉴스
 - 현대미포조선, KCC 지분 1421억원에 처분..주당 35… 머니투데이 | 2016.06.15 | 다음뉴스
 - 현대미포조선, KCC 주식 전량 처분..1652억원 규모 데일리안 | 2016.06.14 | 다음뉴스
- 관련뉴스 8건 전체보기 >

➤ 코스피 시장은 관심을 많이 받았으나 KCC라는 종목은 상대적으로 관심을 덜 받은것으로 보임 (해당 시기에 시장 소외주)

➤ 해당 날짜에 KCC에 관한 기사가 존재하지 않음

```
kospi_data[kospi_data['거래대금'] > 4*(10**13)]
```

	날짜	시가	고가	저가	종가	거래량	거래대금	diff	pct_change
날짜									
2021-01-08	2021-01-08	3040.11	3161.11	3040.11	3152.18	1297903388	409094900005818	0.038386	0.039747
2021-01-11	2021-01-11	3161.90	3266.23	3096.19	3148.45	1712525848	44433772380482	0.054008	-0.001183


```
data.sort_values('거래대금 차이',ascending = False).head(20)
```

	날짜	시가	고가	저가	종가	거래량	거래대금	등락률	diff	pct_change	거래대금 차이
	날짜										
2001-07-31	2001-07-31	63800	64800	63200	64200	15700	1008807000	0.63	0.024922	0.006270	-915384598150
2001-04-16	2001-04-16	46100	46800	46000	46000	4090	188963000	-1.71	0.017391	-0.017094	-917479355050
2001-07-30	2001-07-30	66000	66000	63100	63800	11710	747100000	-1.09	0.045455	-0.010853	-938803418700
2001-04-09	2001-04-09	49900	50700	48750	48750	3890	195359000	-1.52	0.040000	-0.015152	-944342489650
2001-07-16	2001-07-16	71600	74500	71500	73600	12850	947369000	2.94	0.040761	0.029371	-960693280050
2001-08-10	2001-08-10	71400	71500	68800	68900	15040	1049594000	-0.29	0.039187	-0.002894	-988581617800
2001-07-23	2001-07-23	67000	69000	66500	67800	12240	823445000	1.65	0.036873	0.016492	-994794371500
2001-07-05	2001-07-05	72000	72100	69600	72000	13240	943533000	1.41	0.034722	0.014085	-1057599910300
2001-07-10	2001-07-10	70200	73900	70200	73000	20380	1473716000	2.67	0.050685	0.026723	-1078317841500
2003-01-28	2003-01-28	112000	114000	111000	113500	18100	2050075000	0.89	0.026432	0.008889	-1088864459255
2001-07-26	2001-07-26	65700	65700	63600	65000	16060	1035347000	-0.61	0.032308	-0.006116	-1094142445050
2001-09-03	2001-09-03	66200	67500	66100	66500	12080	806156000	0.76	0.021053	0.007576	-1095742945550
2001-10-30	2001-10-30	88000	88000	85500	85500	4170	358980000	-2.84	0.029240	-0.028409	-1104765349200
2001-04-02	2001-04-02	49400	49400	47600	48750	4670	226943000	-0.51	0.036923	-0.005102	-1105112273550
2003-03-10	2003-03-10	86500	88100	86500	86600	21145	1840395000	-0.46	0.018476	-0.004598	-1129178895455
2001-07-11	2001-07-11	72400	75700	71200	74500	26210	1940185000	2.05	0.060403	0.020548	-1135113360350
2001-07-20	2001-07-20	69200	69200	66000	66700	21030	1411559000	-3.33	0.047976	-0.033333	-1135565943200
2001-07-06	2001-07-06	71500	72200	70200	71800	8460	602603000	-0.28	0.027855	-0.002778	-1143001635650
2003-02-03	2003-02-03	112000	113500	111000	112000	21660	2428840000	0.00	0.022321	0.000000	-1148015439625
2001-10-22	2001-10-22	76100	77800	75600	77000	7710	593639000	0.00	0.028571	0.000000	-1153617595750

```
data.sort_values('거래대금 차이',ascending = False).tail(20)
```

	날짜	시가	고가	저가	종가	거래량	거래대금	등락률	diff	pct_change	거래대금 차이
	날짜										
2021-02-03	2021-02-03	204000	206000	200000	206000	41931	8521920000	1.23	0.029126	0.012285	-22419260301837
2021-05-12	2021-05-12	368500	377000	344500	353500	99693	35459136000	-4.07	0.091938	-0.040706	-22422606909899
2021-01-28	2021-01-28	206500	211500	204000	206000	48940	10085354500	-2.60	0.036408	-0.026005	-229977797425890
2021-01-15	2021-01-15	206000	207000	200500	202500	36598	7443104000	-1.22	0.032099	-0.012195	-23154848723562
2021-01-18	2021-01-18	202000	202500	195000	196500	63206	12492930500	-2.96	0.038168	-0.029630	-23174391403611
2021-01-14	2021-01-14	199500	208000	199000	205000	73307	14994922000	2.76	0.043902	0.027569	-24107961668728
2021-01-19	2021-01-19	198500	204000	197000	201000	56045	11240617500	2.29	0.034826	0.022901	-24179212361125
2020-12-28	2020-12-28	201500	204500	193500	194000	89216	17523051500	-2.76	0.056701	-0.027569	-24180257518419
2021-01-22	2021-01-22	212500	221500	212000	215000	92105	19952547000	2.87	0.044186	0.028708	-24285402960985
2021-01-29	2021-01-29	208000	208500	193500	198500	76148	15124007000	-3.64	0.075567	-0.036408	-24335183095379
2021-01-20	2021-01-20	202500	207000	200000	206000	55323	11248589000	2.49	0.033981	0.024876	-24616808269907
2021-01-13	2021-01-13	204500	204500	198000	199500	60402	12084266000	-1.24	0.032581	-0.012376	-24713505920104
2021-01-04	2021-01-04	197500	202000	194000	199000	76218	15136313000	0.76	0.040201	0.007595	-24996257647858
2021-01-21	2021-01-21	208500	210000	205500	209000	38780	8070492000	1.46	0.021531	0.014563	-25271894157736
2021-01-05	2021-01-05	199000	207500	199000	204500	84213	17163099000	2.76	0.041565	0.027638	-26531217080493
2021-01-07	2021-01-07	203000	208000	202500	205500	68550	14082291500	1.48	0.026764	0.014815	-27168725043412
2021-01-06	2021-01-06	208500	210000	202000	202500	81838	16776711500	-0.98	0.039506	-0.009780	-29892619731930
2021-01-12	2021-01-12	203000	204000	198500	202000	63141	12671609500	-1.46	0.027228	-0.014634	-31394725439272
2021-01-08	2021-01-08	207000	210500	204000	210500	87644	18200887500	2.43	0.030879	0.024331	-40891289118318
2021-01-11	2021-01-11	211500	213000	199000	205000	113932	23496661500	-2.61	0.068293	-0.026128	-44410275718982

➤ 단순히 종목거래대금과 코스피 거래대금의 차이값으로 거래데이터 정렬 했을 때 2000년 초반에는 코스피 거래대금이 적고 2020년도에는 코스피거래대금이 커서 차이가 적은값들은 2000년 초반에 몰려있으며 차이가 큰 값들은 2021년에 몰려있으므로 차이를 비교로 사용하는 것은 적절하지 않아보입니다.

➤ 시장 거래대금과 종목거래대금 사이 비교를 할 수있는 지표에 대한 선정이 필요합니다.

코스피 수익률과 개별종목 수익률

➤ 코스피 수익률 과 종목 수익률

Random Sampling을 통해 코스피 200으로부터 30개 종목 sampling을 30번 반복했을 때

코스피 일일수익률과 종목 일일수익률 사이 샘플 상관계수 평균들의 상관계수 평균 0.6953697887386756

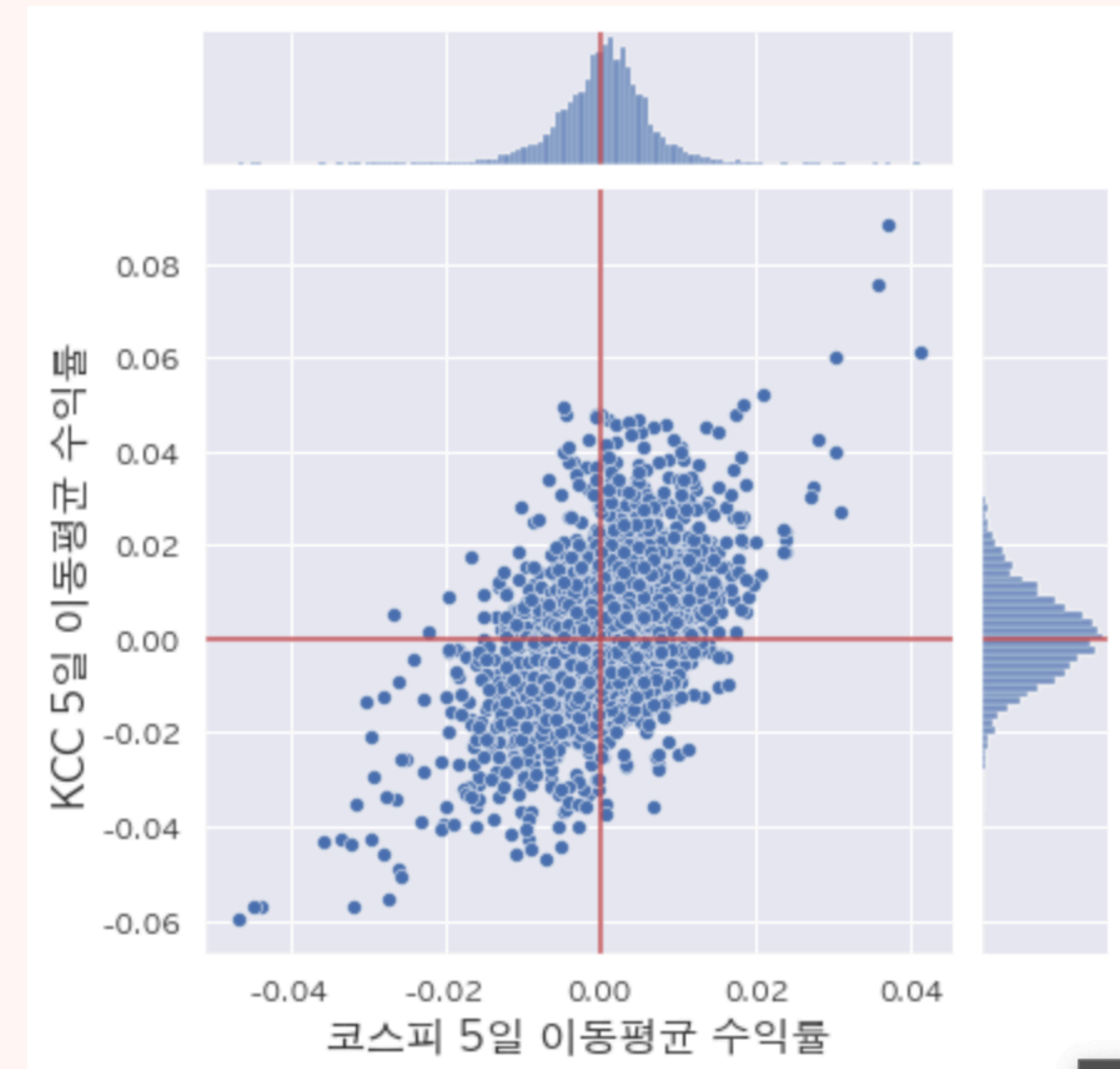
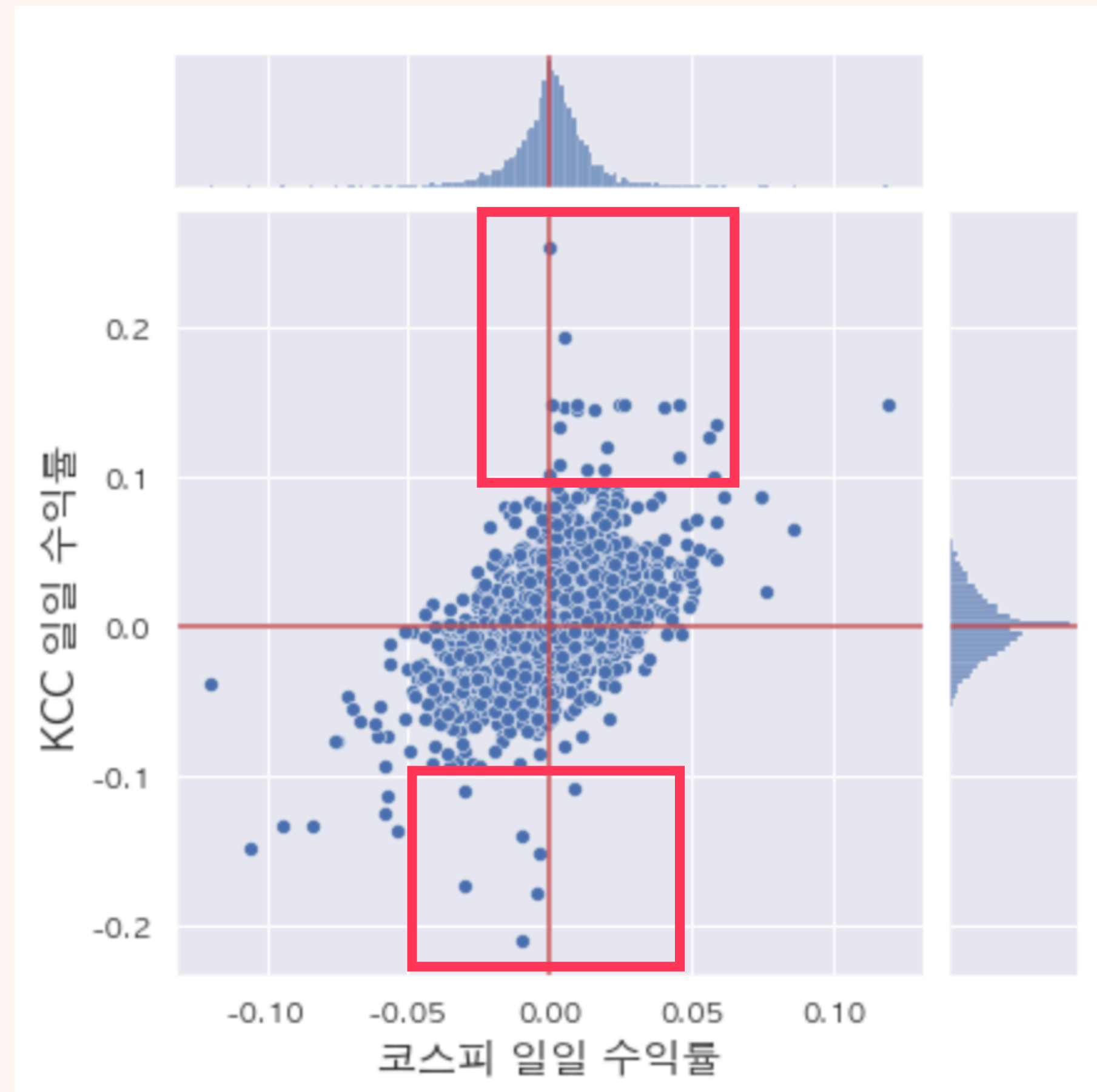
코스피 5일 수익률 이동평균과 종목 5일 수익률 이동평균 사이 샘플 상관계수 평균들의 상관계수 평균 0.7058779775697126

Random Sampling을 통해 코스피 200으로부터 30개 종목 sampling을 100번 반복했을 때

코스피 일일수익률과 종목 일일수익률 사이 샘플 상관계수 평균들의 상관계수 평균 0.6979660664352398

코스피 5일 수익률 이동평균과 종목 5일 수익률 이동평균 사이 샘플 상관계수 평균들의 상관계수 평균 0.7093267906599975

코스피 수익률과 개별종목 수익률



➤ 흐름을 반영하는 코스피 5일 이동평균 수익률과 KCC 5일 이동평균 수익률은 선형성이 강해 직관적으로 그래프에서 어디가 이상점인지 아닌지 판단하기 힘듭니다.

추가 학습

- 시계열 흐름을 활용하여 이상치 탐지 방법이 필요합니다.
- 코스피 200 종목별 뉴스 테이블 따로 구축 합니다.