

5								
#1	안녕하세요 5조 배아프조							
	코로나 충격이 코스피 200에 미치는 영향으로 발표시작하겠습니다.							
#2	저희는 분석배경을 시작으로, ARIMA모델, VAR모델을 가지고 데이터 분석을 하고, 결론을 도출하겠습니다.							
#3	분석 배경							
#4	팀 소개를 하자면 저희 조는 이지현, 전유석, 김정훈, 황윤재, 이명희님이 참여했고, 각자의 관심사, 해커톤 진행간 목표, 그리고 역할은 이러합니다.							
#5	분석 배경							
	지난 중간 발표를 간단하게 요약하자면							
	개인이 주식 증가 예측을 왜 하는가?							
	개인이 주식투자를 하는 이유는 수익							
#6	문제는 각자의 상황을 도외시한 채							
	군중심리에 휩쓸린 투자를 한다는 것							
#7	이를 예방하고자 나름의 투자 기준을 마련해보자가 저희 조의 목표였습니다.							
#8	분석에 앞서 분석의 이론적 틀을 언급하자면							
	기본적 분석이 있고, 기본적 분석에는 경제 분석, 산업군 분석, 개별기업 분석이 존재합니다.							
	위와 같은 분석들을 기반으로 내재가치를 추정하고							
	시장가격과 비교하여 수익을 극대화할 수 있습니다.							
#9	그리고 기술적 분석							
	기술적 분석에는 기본가정 3가지가 있습니다.							
	1. 모든 것은 시장에 반영된다							
	2. 가격은 추세를 이룬다.							

	3. 역사는 반복된다.						
	를 기반으로 저희가 차트를 통해 분석을 할 수 있게 되는 것입니다.						
#10	이 때 저희가 중요하게 본 것은						
	'주가 변동의 원인이 되는 충격'						
	즉, 이슈였습니다.						
	이슈로 인해 주가 변동할 때에 어느 지점에서 포지셔닝을 하느냐						
	급등한다고 따라서 사고						
	떨어진다고 다 파는 것이 아닌						
	기준을 잡아 투자의 안정성을 확보하고자 했습니다.						
#11	그래서 처음 선택한 주제는 우크라이나 전쟁이었습니다.						
	전쟁으로 인한 농산물 ETF의 변화를 보고자						
	유의미한 관계가 있을 것이라고 예측한 원유, 금, 환율, 돼지고기, 방산ETF를 변수로 선정하였습니다.						
#12	하지만 결과는 실패						
	MSE가 20만, R2는 마이너스						
	분석결과 유의미한 결과를 이끌어내지 못했습니다.						
	자체적으로 그 실패이유를 살펴보면						
	짧은 기간이라 데이터가 부족하지 않았나 하는 점과						
	이슈에 대한 변수의 설명력이 부족하지 않았나하는 점이었습니다.						
#13	그래서 재선택한 주제는						
	'코로나 충격이 코스피200에 미치는 영향입니다.'						
	이와 같이 주제를 선택한 이유는						
	앞서 자체적으로 살핀 실패이유를 보완할 수 있기 때문입니다.						
	전쟁보다 코로나 시기가 길기에 보다 많은 데이터로 분석할 수 있고,						
	보다 많은 연구가 이루어졌기에						
	변수 선정에 있어 도움을 받을 수 있는 부분이 있을 것이라 판단했기 때문입니다.						

#14	-						
#15	먼저 ARIMA 모델을 한 문장으로 설명하면						
	과거지식이나 경험을 바탕으로 경제가 움직인다고 간주하고						
	이를 기반으로 끊임없이 훈련하여 좋은 예측치를 제공하는 모델입니다.						
	이는 앞서 전쟁 이슈를 여러 변수로 선정하여 예측하는 것보다						
	좋은 예측이 기대되었기에 모델로 선정했습니다.						
#16	변수 선정						
	코스피 200을 선택한 이유는 시장 대표 지수로, KOSPI에 비해 시장 왜곡 해석가능성이 낮기 때문입니다.						
#17	다음으로 기간 선정						
	코로나 충격이 코스피 200에 미치는 영향을 알아보기 위해						
	2020년 1월 19일 코로나 첫 확진자가 생긴 시점으로부터 2020년 2월 17일, 공포에 물들기 전까지 단기 훈련						
	2020년 1월 1일부터 2020년 2월 17일까지 코로나 충격이전 안정적인 주가를 훈련하여 중기 훈련						
	2019년 1월 1일부터 2020년 2월 17일까지 장기 훈련하여						
	ARIMA 모델이 훈련기간에 따라 어떤 예측치를 보일지 살펴보고자 했고,						
	2020년 2월 18일부터 2020년 3월 19일 공포가 최고조로 달했던 시점을 단기 예측						
	2020년 2월 18일부터 2020년 7월 31일 공포에서 벗어나 원상으로 복귀했던 시점을 중기 예측						
	2020년 2월 18일부터 2022년 4월 30일 현재까지 장기 예측						
#18	이 모형그래프로 검정하여 살펴보면 정상성, 정규성, 상관성을 살펴보면 위와 같고						
#19	오토 ARIMA를 돌렸을 때, pdq가 231, 231, 010입니다.						
	그리고 이 훈련이 적절한 훈련인지 알아보기 위한 독립성 검정, 정규성 검정, 분산성 검정 실시하였고, 그 결과는 이러합니다.						
	독립성 검정의 귀무가설은 자기상관성이 없다.						
	정규성 검정의 귀무가설은 정규분포를 따른다.						

	분산성 검정의 귀무가설은 분산이 일정하다.					
	유의수준 <b>0.05</b> 로 모두 충족한 것은 중기 훈련 뿐이지만					
	훈련 기간에 따른 <b>ARIMA</b> 모델이 어떤 결과를 보이는지 보이기 위해 모두 사용하겠습니다.					
#20	그래프를 보시면 초록색 단기, 빨간색 중기, 보라색 장기로					
	기간에 따라 기울기가 완만해짐을 볼 수 있습니다.					
	단중장기 모두 일정 시점까지만 훈련을 한 경우이므로					
	시장에 반영되지 않았던 충격, 코로나 충격을 예측할 수 없었고, 실제와 다른 예측을 제공했습니다.					
#21	다음은 스텝에 따라 훈련을 한 경우입니다.					
	위와 같이 <b>9</b> 가지 경우의 수를 나누어 검정한 결과					
	<b>A</b> 단기훈련 단기에측만 유의미할 뿐 나머지는 검정 결과 무의미합니다.					
	다만, 검정과정 중에 재미있게 본 수치는 <b>E</b> 중기훈련, 중기에측입니다.					
	코로나 이전 일정정도 안정적으로 움직인 주가를 훈련하여					
	코로나 충격으로 이탈된 주가가 다시 회복한 시점을 예측 해보니					
	분산성 검정이 <b>1</b> 에 육박했습니다.					
	즉, 훈련데이터, 코로나 이전 일정정도 안정적으로 움직인 주가와					
	코로나 충격으로 이탈된 주가가 다시 회복한 주가의 분산성이 유사할 때, 이슈로 인한 충격이 소멸되었다고 볼 수 있습니다.					
	사실 저희 분석은 시간관계상 여기서 매듭지었지만					
	이슈들 그룹화하여 예를 들면 코로나, 메르스, 사스, 조류 인플루엔자 등을 전염병으로 그룹화하여					
	유사한 분산성 검정 흐름이 나타난다면, 전염병이 발생하였을 때 이를 기준으로 주식투자를 할 수 있습니다.					
	그리고 보다 세부적으로 들어가서 코스피 <b>200</b> 뿐만 아니라 산업군, 개별기업들의 차트를 분석하는데 사용할 수 있을 것이라 생각합니다.					
#22	<b>ARIMA</b> 모델 예측 그래프는 아시다시피 끊임없이 훈련하기에 주가를 잘 따라갑니다.					
#23	다만, 변동성이 큰 경우 이탈되는 지점들이 있으며 예측력이 낮지만,					
#24	상대적으로 변동성이 작을 때 그 예측력이 높아집니다.					

#25	MAPE로 모델의 성능을 평가를 보면					
	단기와 중기의 기간 차가 적어 유사하게 나왔고,					
	장기 훈련 예측의 경우 변동성이 높은 단기에는 편차가 크게 나왔습니다.					
	다만, 긴 시점을 예측할수록 실제값과의 차이가 줄어드는 모습을 볼 수 있습니다.					
#26	ARIMA 모델이 좋은 예측력을 보였지만					
	이슈가 어떤 지표에 얼마나 영향을 미치는지					
	어떤 지표가 이슈를 잘 설명하는지는 알 수 없었습니다.					
	그러다 찾은 것이 VAR 모델입니다.					
	VAR 모델이란 내외생 변수의 구분없이 적용할 수 있는 다변량 시계열모형입니다.					
	이는 크게 충격반응분석을 통해 동태적 반응을 파악 가능하고					
	분산분해를 통해 각 내생변수의 변동이 전체 변동에 기여하는 상대적 크기를 분석할 수 있습니다.					
	마지막으로 경제이론보다 실제 자료에서 도출된 결과를 분석할 수 있는데 그래인저 인과성 테스트로					
	각 변수간 인과성을 가능해볼 수 있습니다.					
#27	선정한 변수는 KS200, VIX(미국 변동성 지수), 환율, 금이며					
	코로나로 인해 본격적으로 주가가 하락한 시점에서 회복한 시점까지를 분석 기간으로 삼았습니다.					
#28	히트맵과 산점도를 통해 각 변수 간 상관성을 확인해 보고,					
#29	차분을 하여 추세를 제거하고					
#30	VAR 모형에 넣어 분석하면 이런 결과가 나옵니다.					
#31	이를 그래프로 풀면					
	변동성을 예측하고 그 구간을 가능할 수 있으며					
	각 변수간 얼마큼의 영향이 있는지를 살펴볼 수 있습니다.					
#32	그리고 그래인저 인과성 테스트, 저희가 VAR 모델을 선택한 이유입니다.					

	이를 적용해보면						
#33	이러한 결과가 나옵니다.						
	설명을 하자면 그래인저 인과성 테스트의 귀무가설은 양 변수간 영향력이 없다입니다.						
	VIX가 KS200으로 가는 방향성은 유의수준 0.05 미만이므로						
	시차 1에서 4까지 영향이 있다. 인과성이 있을 수 있다고 해석할 수 있고,						
	그에 역 KS200에서 VIX로 가는 방향성은 유의수준을 넘으므로						
	영향력이 없다고 볼 수 있습니다.						
	이를 통해 계량화하기 어려운 이슈를 몇가지 지표로 인과성을 추정할 수 있으며						
	여기에는 언급하지 않았지만 분산분해를 통해 전체 변동에 대한 상대적 크기 분석하여						
	이슈를 숫자로 만들 수 있지 않을까라는 생각도 해보았습니다.						
#34	결론						
#35	ARIMA 모델의 경우 장기, 중기, 단기로 훈련과 예측을 진행했는데						
	스텝별 훈련을 진행하지 않은 경우 과거 추세에 경향성 파악 정도만 가능하고 예측에는 부적절						
	스텝별 훈련을 진행한 경우 보다 데이터에 민감하게 반응하는 단기 훈련 예측이 예측에 유용						
	VAR 모델의 경우 보다 다양한 변수를 활용할 수 있으며						
	보다 다양한 함수를 통해 다른 관점에서 많은 것을 설명할 수 있다.						
#36	이번 해커톤 간 한가지 목표, 주식투자의 기준을 마련하는 것을 위해						
	크게 두 가지를 살펴보고자 했습니다						
	첫번째는 이슈를 어떤 지표로 설명할 수 있는가						
	두번째는 이슈로 인한 어떤 충격이 발생하였을 때 어떤 차트모양이 그려지는가, 그리고 반복되는가						
	이 둘을 기반으로 기준을 삼아 주식 투자를 한다면 보다 좋은 수익을 있다라고 생각했습니다.						
	문제는 언제 어디서 어떤 이슈가 터질지 모르니 불확실성이 높고						

	한 번 터진 이슈가 반드시 동일한 변동성을 가져올지 가능하기 어려웠습니다.						
	그래도 <b>ARIMA</b> 모델에서 <b>E</b> (중기 훈련, 중기 예측)에서 나온 분산성						
	분산성 검정을 통해 충격을 해소 정도를 예측해 볼 수 있지 않을까와						
	<b>VAR</b> 모델에서 인과성을 설명할 수 있는 그래인저 인과성 테스트						
	이를 기반으로 이슈를 잘 설명하는 변수를 찾을 수 있지 않을까란 생각을 하게 되었습니다.						
	다음에 기회가 있다면 딥러닝을 통해 보다 많은 변수들을 다루고,						
	이슈를 보다 정확히 해석하고 주식 투자에 있어 기준을 마련하고 싶습니다.						
#37	마지막으로 프로젝트 평가를 위해 완성도 비교표를 그렸습니다.						
	이번 해커톤에서 저희 조가 기대했던 바와 자체 평가를 평균낸 것으로 달성도를 살펴보면 이릅니다.						
#38	이상으로 발표 마치겠습니다.						
	질문 있으신가요?						