

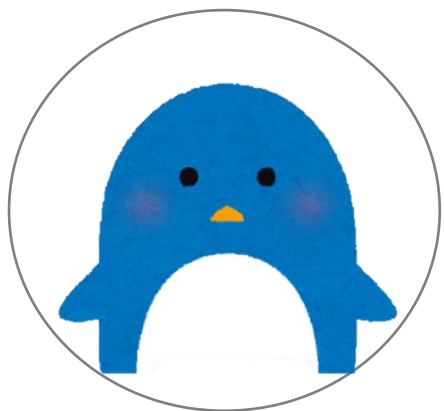
2집 ★맛집★ 2조

포스트 코로나 시대에 대응하는 국내 수출입 품목별 경쟁력 예측

- 자동차부품, 반도체 품목에 대한 수출입 예측 모델 수립을 중심으로 -

INTRODUCE

팀원 및 역할분담



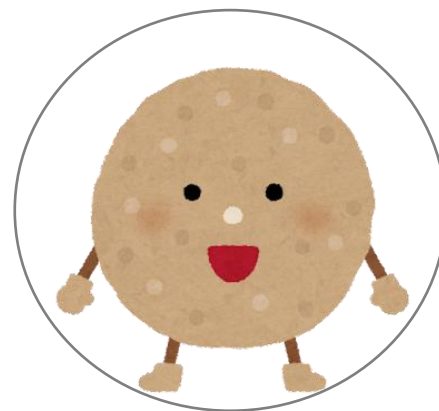
김현아

- 자료조사 및 데이터 수집
- 데이터 전처리
- 반도체 수출입 ARIMA
- 반도체 수출입 다중회귀
- PPT 제작, 발표



박혜인

- 자료조사 및 데이터 수집
- 데이터 전처리
- 자동차부품 수출입 ARIMA
- 자동차부품 수출입 다중회귀
- PPT 제작, 최종 수정



임승찬(조장)

- 자료조사 및 데이터 수집
- 데이터 전처리 (TSI,RCA)
- TSI 지수 ARIMA 분석
- PPT 제작



조완제

- 자료조사 및 데이터 수집
- 데이터 전처리(TSI,RCA)
- TSI 지수 ARIMA 분석
- PPT 제작

CONTENTS

01	프로젝트 기획 배경 및 목표	004	03	결론 및 기대효과	055
	<ul style="list-style-type: none">- 기획배경 및 목표- 주제에 대한 설명 및 품목 선정			<ul style="list-style-type: none">- 모델링을 통한 최종적인 결론- 개선사항 및 기대효과	
02	데이터 분석 시각화 및 모델링	011	04	부록	058
	<ul style="list-style-type: none">- 다중회귀분석- ARIMA 시계열 분석			<ul style="list-style-type: none">- 모델링 후기 및 느낀점- 참고문헌	

Part 1

프로젝트 기획 배경 및 목표

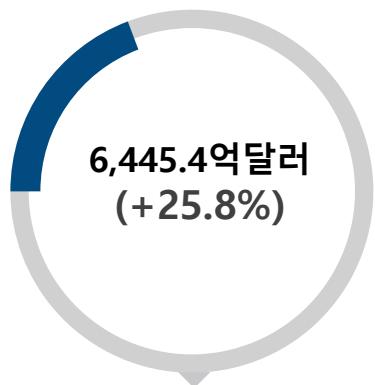
- 기획배경 및 목표
- 주제에 대한 설명 및 품목 선정

프로젝트 기획 배경 및 목표

01 프로젝트 기획 배경

① 국내 전통 산업의 견고한 성장

2021년 국내 수출액



글로벌 무역 순위



출처: 산업통상자원부 수출입 동향

- 반도체 산업 자동차산업, 등과 같은 전통 산업의 견고한 성장으로 인한 21년 역대 **최대 무역 실적 달성**

② ICT 기술 발달에 따라 반도체 및 자동차 부품 수요 증가



- 코로나 19 확산 이후 디지털 혁명이 더욱 발전하는 중이며 최근 반도체 및 자동차 부품 등과 같은 디지털 관련 **주요 수출 품목 수요 증가**

프로젝트 기획 배경 및 목표

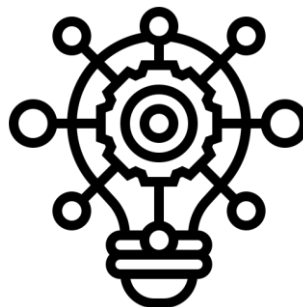
02 프로젝트 목표 (4가지)



포스트 코로나 시대의
자동차부품, 반도체 품목
수출입 예측 모델 수립



한국과 주요국 간의 품목별
수출경쟁력 비교 분석 및 예측



변화하는 수출입에 따른
품목별 공급 및 수요 시사점에
대한 인사이트 도출



국내 자동차 부품과 반도체
산업의 경쟁력 제고에
필요한 기초 자료 제공

주제에 대한 설명 및 품목 선정

01 주제에서 말하는 '경쟁력'의 의미

(1) 품목별 경쟁력이란 ?

현시비교우위지수(RCA)와 무역특화지수(TSI)를 통해 선정한 품목이 다른 국가와 비교하였을 때, 우리나라 수출품목의 우열을 가리게 하는 힘을 의미함

(2) 무역특화지수(TSI) 란?

국가(i)가 세계(j)에 수출하는 품목(k)의 량과 수입하는 량을 이용하여 계산하는 지수
 $1 < \text{TSI 지수} < 1$, 1에 가까울 수록 수출특화, -1에 가까울수록 수입특화



$$TSI_{ij}^k = \frac{EX_{ij}^k - IM_{ij}^k}{EX_{ij}^k + IM_{ij}^k}$$

(3) 현시비교우위지수(RCA) 란?

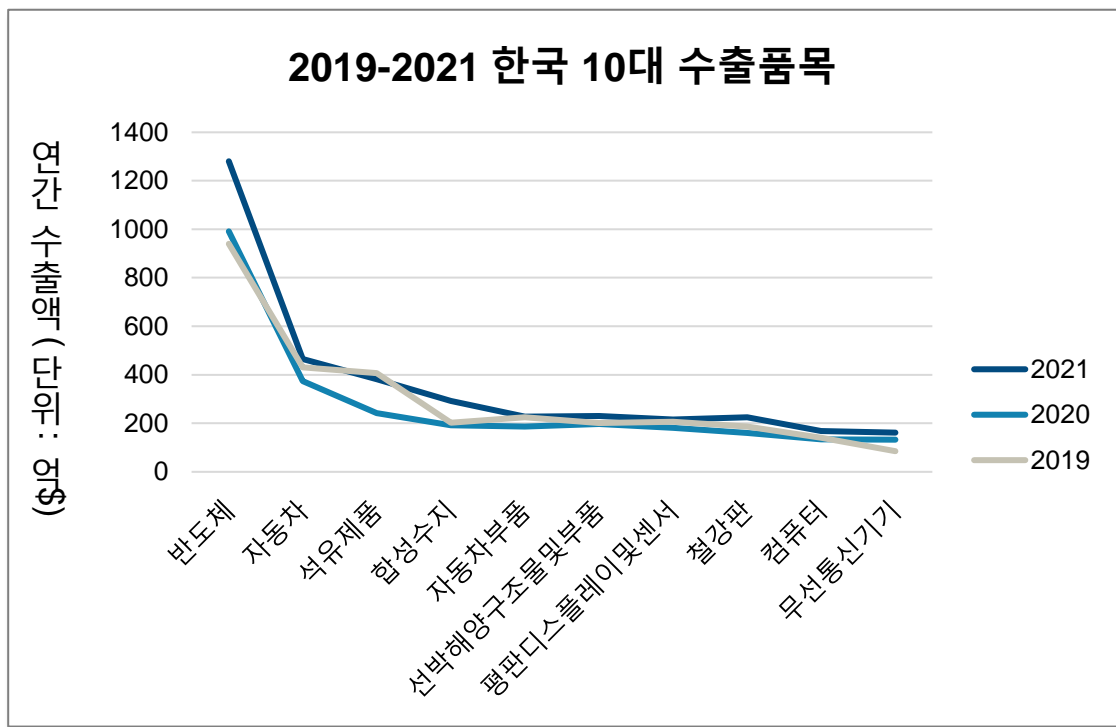
수출하는 품목(k)이 자국(i)의 다른 품목에 비해서 얼마나 수출경쟁력을 가지고 있는지 판단
 $RCA > 1$ 자국의 다른 품목에 비해서 주로 수출되는 품목이다



$$RCA_i^k = \frac{EX_i^k}{EX_W^k} / \frac{EX_i}{EX_W}$$

주제에 대한 설명 및 품목 선정

02 품목 선정 이유



- 한국의 10대 수출품목 중 연간 수출액이 가장 높은 **반도체**와 수출액 5위인 **자동차부품** 품목을 선택하였음
- 코로나19 백신의 보급 이후로 '**온택트**' **트렌드**가 보편화 되어 반도체 산업이 더욱 증가 할 것으로 생각되어 선택하였음
- 현재 국내에서 자동차 부품 산업의 수출이 급감. 부품 수급 차질로 인한 완성차 **생산공장의 가동중단**과 **글로벌 완성차 판매 감소** 때문
- 코로나 19의 영향으로 생산/판매가 감소했는데, 과연 코로나 종식 후에는 증가할 것인지 예측해보기 위해 자동차 부품 산업을 선택함

주제에 대한 설명 및 품목 선정

03 품목별 피처 변수 선정 과정

(1) 특성 선별 과정

변수이름 (반도체)	변수이름 (자동차부품)
글로벌 금융위기 변동(0,1)	해외투자 유치 유무(0,1)
유가지수	세계경제성장률
물가상승률	상하이 컨테이너운임지수(SCFI)
금리인상	고용률
항공운임	공장가동률
신규생산라인가동	부품생산량
관련법여부	자동차부품산업의 가격하락
반도체품질	중소기업비율의 변화
반도체가격	전기차 생산량(친환경자동차)
산업자원부의 기술수준비교한 수치	수출편향지수EBI
무역수지	ECI
원자재가격	기업체수
세계경제성장률	종업원(정규직숫자)
환율	ESI
실질실효환율	TSI
IIT	RCA
GDP대비 무역의존도	환율
감응도계수	실질실효환율
영향력계수	품질경쟁력
생산유발계수	기술경쟁력()
부가가치 유발계수	가격경쟁력
ESI	서비스경쟁력
TSI	변수이름 (자동차부품)
RCA	해외투자 유치 유무(0,1)
ECI	세계경제성장률

- 반도체 산업, 자동차부품 산업에 대해 자료 조사를 시작
- 각각 조사한 자료를 바탕으로 수입 및 수출에 영향을 줄 수 있을 것으로 판단되는 특성을 제안하고, 설명하는 과정을 통해서 약 25개의 특성을 선택
- 통계청, 무역협회, 관세청, 공공정보 데이터, 선행연구논문, 뉴스보도 자료 등의 자료조사를 진행하였음

(2) 최종 피처 선정

- 각 특성을 나타내는 Raw data 조사를 시작
- 자료수집이 불가능한 특성, 시기가 맞지않는 특성 등을 제거하고, 총 8개의 특성으로 결정하였음

특성	설명
환율	월 단위 원/\$ 환율
소비자물가지수	월 단위 소비자 물가지수 (2020=100)
기준금리	한국은행 기준금리
국제유가	월 단위 Dubai유 기준 가격
SCFI	월 단위 상하이 컨테이너운임지수
항공운임	월 단위 푸동 기준
고용률	월 단위 만 15세 이상 인구 중 취업자가 차지하는 비율
가동률지수	월 단위 생산능력 대비 생산실적의 비율

Part 1

Work Breakdown Structure

Gantt Chart

항목	내용	소요 기간 (일)	2022년 03월														
			17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
분석 기획	프로젝트 이해 및 범위 설정	3										Pros 시험					
	프로젝트 정의 및 계획 수립	2															
데이터준비	필요데이터 정의	1															
	데이터 수집 및 크롤링 적절성 점검	2															
데이터분석	데이터 전처리	1															
	분석용 데이터 준비	1															
	EDA	2															
	모델링	4															
	모델 평가 및 검증	4															
평가 및 전개	모델 발전계획 수립	1															
	프로젝트 결과 도출 및 제언	2															
비고		15			주말							주말					

Part 2

데이터 분석 시각화 및 모델링

- 사용 데이터 목록 / 전처리
- 다중회귀분석
- ARIMA 시계열 분석

Part 2-1

사용 데이터 목록 및 전처리

- 사용 데이터 목록
- 데이터 전처리

Part 2

사용 데이터 목록

출처	자료원 명	시기	산출지표
ITC(International Trade Center)	품목별 수출입 : 반도체	2011.01~2021.11	TSI, RCA, 시계열 분석 ARIMA
	품목별 수출입 : 자동차 부품	2011.01~2021.11	
한국무역협회	품목 수출입(MTI 742, MTI 831)	2011.01~2022.02	다중회귀 모델 종속변수, ARIMA
	품목 수출입(HS 6단위 : 870810 외 11개)	2011.01~2021.11	TSI, RCA, 시계열 분석 ARIMA
	품목 수출입(HS 6단위 : 847330 외 17개)	2011.01~2021.11	
KOSIS	고용률(시도)	2015.01~2022.02	다중회귀 모델 독립변수
	제조업 생산능력 및 가동률지수(2015=100) 중 반도체 제조업		
	제조업 생산능력 및 가동률지수(2015=100) 중 자동차 부품 제조업		
	소비자물가지수(2020=100)		
	환율		
TRADLINX	상하이 컨테이너 운임지수		
한국관세물류협회	항공운임(인천공항 기점)		
한국석유공사 Petronet	일일국제원유가격		
한국은행 결제통계시스템	한국은행 기준금리		

데이터 전처리

01 회귀분석용 데이터셋 구성 (자동차부품, 반도체)

(1) 자동차 부품 회귀분석용 데이터 셋 구성

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	국제유가	SCFI	고용률(%)	가동률지수(원지수)
0	2015.01	2262441	410713	1,093.50	94.643	2.00	45.77	1051.18	59	105.8
1	2015.02	1963200	347320	1,098.40	94.587	2.00	55.69	1071.73	59.1	90.5
2	2015.03	2371707	459833	1,109.50	94.596	1.75	54.69	868.47	59.8	107.8
3	2015.04	2493058	464552	1,072.40	94.625	1.75	57.72	748.39	60.6	106.7
4	2015.05	2168130	424571	1,108.20	94.890	1.75	63.02	731.35	61.2	97.2
...
79	2021.08	1579116	507368	1,159.50	102.750	0.75	69.50	4308.30	61.2	77.2
80	2021.09	1853552	500444	1,184.00	103.170	0.75	72.63	4590.24	61.3	78.3
81	2021.1	1784922	495623	1,168.60	103.350	0.75	81.61	4596.59	61.4	85.5
82	2021.11	1841532	477131	1,187.90	103.870	1.00	80.30	4561.78	61.5	93.9
83	2021.12	2120744	515085	1,188.80	104.040	1.00	73.21	4887.07	60.4	98.8

84 rows × 10 columns

- 수집한 csv파일 중 필요한 데이터만 남기고 각 피처의 이름으로 데이터 프레임 형태로 저장
- 마지막으로 concat()을 사용해 하나의 데이터 프레임으로 생성

(2) 반도체 회귀분석용 데이터 셋 구성

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	국제유가	항공 운임	고용률(%)	가동률지수(원지수)
0	2015.01	5262694	3353905	1,093.50	94.643	2	45.77	1.226	59	79.7
1	2015.02	4760336	2822967	1,098.40	94.587	2	55.69	1.226	59.1	78.9
2	2015.03	5388654	3199381	1,109.50	94.596	1.75	54.69	1.128	59.8	88.7
3	2015.04	5162367	3115806	1,072.40	94.625	1.75	57.72	1.038	60.6	90.1
4	2015.05	5111415	2892591	1,108.20	94.890	1.75	63.02	1.058	61.2	92.2
...
79	2021.08	11695023	5311582	1,159.50	102.750	0.75	69.50	3.160	61.2	132.3
80	2021.09	12180206	5458903	1,184.00	103.170	0.75	72.63	3.160	61.3	130.8
81	2021.10	11173464	5410830	1,168.60	103.350	0.75	81.61	2.740	61.4	127.9
82	2021.11	12035996	6046221	1,187.90	103.870	1	80.30	2.680	61.5	127.8
83	2021.12	12780662	6356192	1,188.80	104.040	1	73.21	2.580	60.4	136.3

84 rows × 10 columns

- 수집한 csv파일 중 필요한 데이터만 남기고 각 피처의 이름으로 데이터 프레임 형태로 저장
- 마지막으로 concat()을 사용해 하나의 데이터 프레임으로 생성

데이터 전처리

02 시계열 분석용 데이터셋 구성 (자동차부품, 반도체)

(1) 자동차 부품 시계열 분석용 데이터 셋 구성

	년월	수출 금액	수출 증감률	수입 금액	수입 증감률
2	2011.01	1770908	-6.9	521846	3.9
3	2011.02	1697007	-4.2	447911	-14.2
4	2011.03	2044642	20.5	597860	33.5
5	2011.04	1984151	-3	526234	-12
6	2011.05	1882387	-5.1	547208	4
...
129	2021.08	1579116	-20.7	507368	-6.9
130	2021.09	1853552	17.4	500444	-1.4
131	2021.10	1784922	-3.7	495623	-1
132	2021.11	1841532	3.2	477131	-3.7
133	2021.12	2120744	15.2	515085	8

132 rows × 5 columns



	시점	수출액	수입액
0	2011-01-31	1770908	521846
1	2011-02-28	1697007	447911
2	2011-03-31	2044642	597860
3	2011-04-30	1984151	526234
4	2011-05-31	1882387	547208
...
127	2021-08-31	1579116	507368
128	2021-09-30	1853552	500444
129	2021-10-31	1784922	495623
130	2021-11-30	1841532	477131
131	2021-12-31	2120744	515085

132 rows × 3 columns

- 시계열 분석에 필요한 수출액과 수입액으로 새 데이터 프레임 생성
- 시계열 분석을 위해 시점 값을 가장 마지막 날로 지정
date_range('2011-1-1', '2021-12-31', freq='M')

(2) 반도체 시계열 분석용 데이터 셋 구성

	시점	수출 금액	수출 증감률	수입 금액	수입 증감률
0	2011.01	20373544	5.1	8445306	0.9
1	2011.02	16665092	-18.2	6262048	-25.9
2	2011.03	22197493	33.2	8478018	35.4
3	2011.04	19750481	-11.0	7857524	-7.3
4	2011.05	18849229	-4.6	7487074	-4.7
...
127	2021.08	15057897	-15.6	4499001	18.4
128	2021.09	15752184	4.6	4364291	-3.0
129	2021.10	17222409	9.3	4201917	-3.7
130	2021.11	16940429	-1.6	4587520	9.2
131	2021.12	16461673	-2.8	5610036	22.3

132 rows × 5 columns



	시점	수출액	수입액
0	2011-01-31	20373544	8445306
1	2011-02-28	16665092	6262048
2	2011-03-31	22197493	8478018
3	2011-04-30	19750481	7857524
4	2011-05-31	18849229	7487074
...
127	2021-08-31	15057897	4499001
128	2021-09-30	15752184	4364291
129	2021-10-31	17222409	4201917
130	2021-11-30	16940429	4587520
131	2021-12-31	16461673	5610036

132 rows × 3 columns

- 시계열 분석에 필요한 수출액과 수입액으로 새 데이터 프레임 생성
- 시계열 분석을 위해 시점 값을 가장 마지막 날로 지정
date_range('2011-1-1', '2021-12-31', freq='M')

데이터 전처리

03 자동차부품, 반도체 TSI 지수 계산

• $TSI_{ij}^k = \frac{EX_{ij}^k - IM_{ij}^k}{EX_{ij}^k + IM_{ij}^k}$  iloc을 컬럼을 가져와 위 수식처럼 계산

(1) 자동차 부품 TSI

	시점	완충기	안전벨트	기타	기어박스	자동차장치 및 드라이브 축	로트휠	서스펜션	라디에이터	소음기와 배기기관	클러치	운전대	부품품
0	2011.01	0.462985	0.211900	0.690162	-0.205413	0.293579	0.137571	0.454618	0.358986	0.471508	0.503850	0.260434	0.832795
1	2011.02	0.305730	0.045597	0.718448	-0.159416	0.332062	0.326775	0.529058	0.328404	0.478939	0.529941	0.273117	0.860857
2	2011.03	0.377142	0.165180	0.707325	-0.116089	0.291593	0.251942	0.529535	0.173879	0.505488	0.495672	0.202941	0.844138
3	2011.04	0.209599	0.164351	0.711140	-0.037595	0.553601	0.339561	0.475617	0.525808	0.547201	0.506741	0.171547	0.853058
4	2011.05	0.361431	0.455107	0.773034	-0.114981	0.485881	0.355225	0.559879	0.189098	0.449973	0.534950	0.243874	0.833659
...
127	2021.08	-0.527672	-0.041105	0.609070	0.300472	0.737439	0.079129	0.531680	0.173414	0.429498	0.632449	0.461214	0.806707
128	2021.09	-0.530570	0.348868	0.655534	0.438857	0.780009	0.185359	0.562974	0.219067	0.365692	0.682337	0.474773	0.833068
129	2021.10	-0.508787	0.169835	0.640584	0.476016	0.840402	0.105003	0.560311	0.349295	0.423633	0.683637	0.521768	0.811801
130	2021.11	-0.374674	0.106669	0.656894	0.473832	0.843302	0.017573	0.588996	0.442674	0.393873	0.695506	0.538071	0.835817
131	2021.12	-0.364267	0.224990	0.663677	0.555920	0.818046	0.008935	0.581741	0.329670	0.454070	0.687738	0.564809	0.843937

(2) 반도체 TSI

	연도	메모리반도체	시스템반도체	집적회로반도체 부품	트랜지스터	다이오드	기타 개발소자반도체	개별소자반도체 부품	실리콘웨이퍼
0	2011.01	0.452176	-0.591096	0.264422	-0.212914	-0.437548	0.105098	0.722306	-0.316265
1	2011.02	0.464098	-0.554155	0.296639	-0.164408	-0.444315	0.209772	0.722368	-0.286365
2	2011.03	0.512073	-0.500628	0.270785	-0.146302	-0.429300	0.135468	0.720367	-0.274149
3	2011.04	0.519453	-0.479960	0.268016	-0.144329	-0.368342	0.117046	0.702695	-0.306220
4	2011.05	0.494071	-0.479969	0.233814	-0.097801	-0.361180	0.089888	0.704324	-0.364581
...
126	2021.07	0.558269	0.024510	0.224391	0.032528	-0.246395	-0.086086	-0.041393	-0.336198
127	2021.08	0.586470	0.068220	0.268916	-0.006946	-0.334104	-0.077068	-0.000221	-0.376846
128	2021.09	0.586842	0.057215	0.346665	0.045280	-0.318823	-0.233495	-0.129830	-0.345480
129	2021.10	0.542799	0.081229	0.235223	0.015141	-0.311239	-0.160808	-0.015860	-0.363158
130	2021.11	0.497912	0.104803	-0.106693	0.012949	-0.354486	-0.190078	-0.144270	-0.350447

데이터 전처리

03 자동차부품, 반도체 TSI 지수 계산

- $$RCA_i^k = \frac{EX_i^k}{EX_W^k} / \frac{EX_i}{EX_W}$$

 iloc을 컬럼을 가져와 위 수식처럼 계산

(1) 자동차 부품 RCA

	시점	완중기	안전벨트	기타	기어박스	자동차장치 및 트라이브룩	로드휠	서스펜션	라디에이터	소음기와 배기기관	클러치	운전대	부분품
0	2011	0.250638	0.624681	1.044777	0.912599	0.524690	0.726310	0.511356	0.211699	0.963579	0.989737	0.758546	4.969442
1	2012	0.371695	0.971169	1.388344	1.547661	0.762423	0.878288	0.792763	0.375200	1.091926	0.929088	1.072393	4.633702
2	2013	0.367354	1.136122	1.423988	1.687571	0.886924	0.806230	0.952505	0.423111	1.094192	1.097971	1.515339	4.464154
3	2014	0.329087	1.339931	1.549047	1.805815	0.951058	0.889107	0.912596	0.427035	1.004039	1.118463	1.375615	4.170594
4	2015	0.304357	1.433306	1.740747	2.267727	1.091991	1.109497	0.914474	0.411203	0.965589	1.315264	1.397216	4.542987
5	2016	0.295535	1.259304	1.798752	2.420041	1.239252	1.141323	1.007744	0.488354	0.961545	1.908105	1.899614	4.131891
6	2017	0.245420	1.209026	1.151289	1.604565	0.928561	0.880935	0.829855	0.399963	0.523563	1.467022	1.500830	3.115606
7	2018	0.230125	0.916183	1.037489	1.433212	0.921476	0.718868	0.739138	0.377605	0.445826	1.119945	1.412079	2.741898
8	2019	0.249878	1.291234	1.157153	1.491567	1.138990	0.861374	0.797161	0.327087	0.492715	1.369543	1.697870	3.316787
9	2020	0.291835	0.999100	1.135995	1.454255	1.309561	0.721885	0.872984	0.301757	0.366091	1.526929	1.784370	3.422453

(2) 반도체 RCA

	연도별	메모리반도체	시스템반도체	집적회로반도체 부품	트랜지스터	다이오드	기타 개발소자반도체	개별소자반도체 부품	실리콘웨이퍼
0	2011	3.974080	0.214189	0.728112	1.044361	0.683035	1.467136	1.908269	2.346645
1	2012	3.428088	0.318785	0.610085	1.084374	0.803123	2.021191	2.125503	2.809322
2	2013	4.194340	0.365486	0.675208	1.062951	0.812489	2.137392	2.583600	2.800790
3	2014	5.086188	0.342653	0.877327	0.989887	0.633180	1.778695	2.066301	2.279984
4	2015	5.015110	0.262830	0.676521	0.938805	0.562331	1.695310	1.498037	1.880798
5	2016	5.407542	0.278769	0.499574	1.065116	0.606614	2.103084	1.252054	1.781991
6	2017	7.468672	0.283330	6.845199	1.062239	0.705898	2.261636	0.814591	1.629974
7	2018	8.473355	0.292584	5.324552	1.130509	0.823237	2.233798	0.559109	2.763085
8	2019	7.199880	0.344160	0.561585	1.317658	0.894398	1.909834	0.561234	2.390120
9	2020	6.852720	0.399351	0.521820	1.239281	0.761592	1.658356	0.581338	2.204980

Part 2-2

다중회귀분석

- 자동차 부품
- 반도체

다중회귀분석 모델링 과정

01 모듈 및 데이터 불러오기

- 활용 데이터 : 자동차부품, 반도체 데이터셋.csv

02 데이터 전처리

- X : 독립변수에 해당하는 피쳐들
- y : 종속변수에 해당하는 피쳐

03 선형회귀 모델 구축 (OLS)

- ① 선형회귀 모델 가정 확인 : 정규성, 등분산성
- ② Train 데이터의 summary() 값을 확인 후
피쳐 별 p-value값 확인하여 유의하지 않은 피쳐는 삭제
- ③ p-value값이 유의한 피쳐만 가지고 모델 선언 및 학습

04 예측 : 미래의 반응변수 값 예측하기

- 실제 값과 모델 출력 값이 얼마나 유사한지 scatter plot을 활용하여 화면에 표현

05 해당 모델의 성능평가

- 해당 모델의 학습용, 테스트용 데이터에 대하여 평가지표인 MSE, RMSE, MAE, R2_Score 값 확인

06 실제 데이터와 비교

- 완성된 회귀식에 실제 2022년 1~2월의 X값을 대입하여 예측값과 실제값이 얼마나 차이나는지 확인

다중회귀분석 : 자동차부품 수출액

01 활용 데이터 및 전처리

- 활용데이터 : 자동차부품 데이터셋.csv (2015~2021)
- 독립변수 X : 환율, 소비자 물가지수, 기준금리(연%), 국제유가, SCFI, 고용률(%), 가동률지수(원지수)

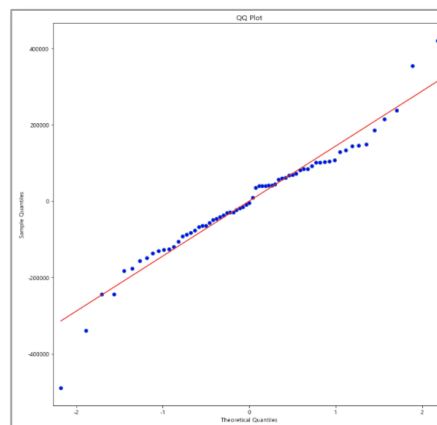
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 84 entries, 0 to 83
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   시점                  84 non-null    float64
1   수출액                84 non-null    int64
2   수입액                84 non-null    int64
3   환율                  84 non-null    float64
4   소비자 물가지수       84 non-null    float64
5   기준금리(연%)         84 non-null    float64
6   국제유가              84 non-null    float64
7   SCFI                  84 non-null    float64
8   고용률(%)             84 non-null    float64
9   가동률지수(원지수)    84 non-null    float64
dtypes: float64(8), int64(2)
memory usage: 6.7 KB
```

Null값 없음
Type도 int, float 형태
→ 추가 인코딩 필요 없음

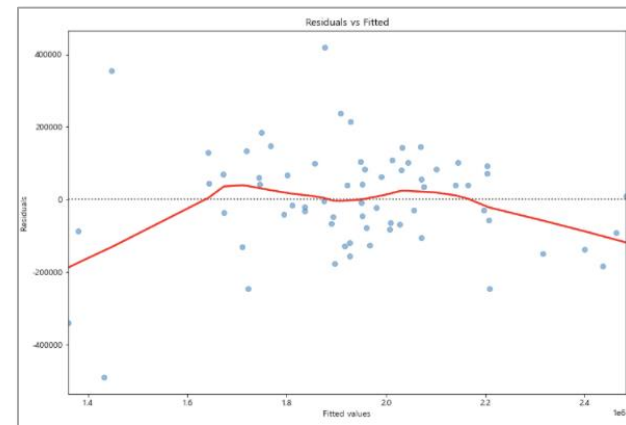
02 선형회귀 모델 구축 (OLS)

- 학습용 데이터셋과 테스트용 데이터셋 분리 (train_test_split)

(1) 선형회귀모델 가정확인



< 정규성 검정 >



< 등분산성 검정 >

- 정규성, 등분산성 모두 만족함을 확인할 수 있음 → **OLS 사용가능**

Part 2-2

다중회귀분석 : 자동차부품 수출액

(2) OLS Regression 결과 확인

OLS Regression Results

Dep. Variable:	수출액	R-squared:	0.725
Model:	OLS	Adj. R-squared:	0.692
Method:	Least Squares	F-statistic:	22.23
Date:	Thu, 31 Mar 2022	Prob (F-statistic):	2.17e-14
Time:	12:40:02	Log-Likelihood:	-891.00
No. Observations:	67	AIC:	1798.
Df Residuals:	59	BIC:	1816.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	7.082e+06	2.27e+06	3.119	0.003	2.54e+06	1.16e+07
환율	-724.6979	568.598	-1.275	0.207	-1862.460	413.064
소비자 물가지수	-6.538e+04	1.91e+04	-3.420	0.001	-1.04e+05	-2.71e+04
기준금리(연%)	3292.9839	8.59e+04	0.038	0.970	-1.69e+05	1.75e+05
국제유가	5604.2952	3178.768	1.763	0.083	-756.405	1.2e+04
SCFI	66.2822	29.924	2.215	0.031	6.404	126.160
고용률(%)	4116.1505	2.29e+04	0.180	0.858	-4.17e+04	5e+04
가동률지수(원지수)	1.628e+04	2214.139	7.352	0.000	1.18e+04	2.07e+04

- 유의수준 0.05를 넘는 p-value를 가진 피처변수 제거 필요
- 제거할 피처: 환율, 기준금리, 고용률(%)

(3) 유의성 검정 후 피처변수 제거 / 모델 선언 및 학습

OLS Regression Results

Dep. Variable:	수출액	R-squared:	0.716
Model:	OLS	Adj. R-squared:	0.698
Method:	Least Squares	F-statistic:	39.17
Date:	Thu, 31 Mar 2022	Prob (F-statistic):	2.49e-16
Time:	12:40:10	Log-Likelihood:	-892.03
No. Observations:	67	AIC:	1794.
Df Residuals:	62	BIC:	1805.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	7.27e+06	1.36e+06	5.356	0.000	4.56e+06	9.98e+06
소비자 물가지수	-7.383e+04	1.33e+04	-5.535	0.000	-1e+05	-4.72e+04
국제유가	7129.5778	1701.588	4.190	0.000	3728.151	1.05e+04
SCFI	70.6213	26.636	2.651	0.010	17.377	123.865
가동률지수(원지수)	1.606e+04	2178.844	7.369	0.000	1.17e+04	2.04e+04

가

3개의 변수

회귀계수 확인가능

3개의 피처 제거 후 나머지 변수에 대해 유의함을 확인

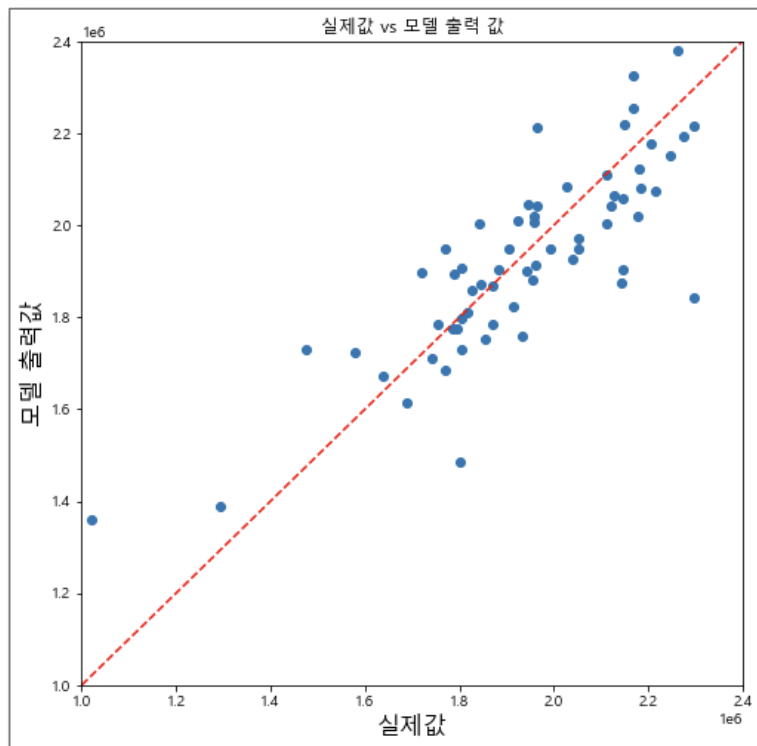
➤ 회귀식 도출

$$\text{수출액} = -7.383e+04 * \text{소비자 물가지수} + 71.29.5778 * \text{국제유가} + 70.6213 * \text{SCFI} + 1.606e+04 * \text{가동률지수(원지수)} + 7.27e+06$$

다중회귀분석 : 자동차부품 수출액

03 예측 : 미래의 반응변수 값 예측하기

- 실제값과 모델 출력값이 거의 비슷함을 예측할 수 있음



04 해당 모델의 성능 평가

- MSE, RMSE, MAE, R2_Score 값 확인

테스트용 데이터셋 결과
MSE: 25735000074.895
RMSE: 160421.321
MAE: 113889.055
R2_Score: 0.818

- ▶ 테스트용 데이터셋 성능평가 결과, 결정계수가 81.8%의 정확도
- ▶ 각각 **RMSE가 160,421**이므로, **예측값과 실제값의 차이가 이 수치만큼 차이날 수 있음**을 예상할 수 있음

다중회귀분석 : 자동차부품 수출액

05 예측치와 실제 데이터 비교

- 회귀분석에 사용되지 않은 2022년 1월, 2월 데이터셋을 활용하여 실제값과 비교

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	항공운임	가동률지수(원지수)	고용률
0	2022.01	10816657	5739959	1205.50	104.69	1.25	2.79	122.9	59.6
1	2022.02	10381788	5493588	1197.83	105.30	1.25	2.85	122.6	60.6

- 위에서 산출한 회귀식에 해당 독립변수의 값들을 넣어 값을 예측

$$\begin{aligned} \text{수출액} = & -7.383\text{e}+04 * \text{소비자 물가지수} + 71.29.5778 * \text{국제유가} \\ & + 70.6213 * \text{SCFI} + 1.606\text{e}+04 * \text{가동률지수(원지수)} + 7.27\text{e}+06 \end{aligned}$$

- 예측 결과

예측 2022년 1월 수출액: 1951919
실제 2022년 1월 수출액: 2061375

예측 2022년 2월 수출액: 1906552
실제 2022년 2월 수출액: 1783051



- 2022년 1~2월 자동차부품 수출액은 예측값과 실제값의 차이가 RMSE 값인 160,421와 유사하게 차이가 남
- 수출액을 거의 유사하게 예측함을 확인할 수 있음

다중회귀분석 : 자동차부품 수입액

01 활용 데이터 및 전처리

- 활용데이터 : 자동차부품 데이터셋.csv (2015~2021)
- 독립변수 X : 환율, 소비자 물가지수, 기준금리(연%), 국제유가, SCFI, 고용률(%), 가동률지수(원지수)

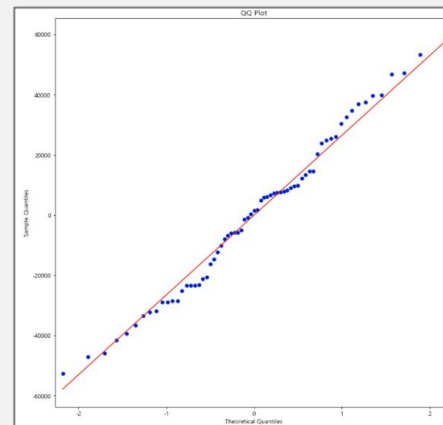
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 84 entries, 0 to 83
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   시점                  84 non-null    float64
1   수출액                84 non-null    int64  
2   수입액                84 non-null    int64  
3   환율                  84 non-null    float64
4   소비자 물가지수       84 non-null    float64
5   기준금리(연%)         84 non-null    float64
6   국제유가              84 non-null    float64
7   SCFI                  84 non-null    float64
8   고용률(%)            84 non-null    float64
9   가동률지수(원지수)    84 non-null    float64
dtypes: float64(8), int64(2)
memory usage: 6.7 KB
```

Null값 없음
Type도 int, float 형태
→ 추가 인코딩 필요 없음

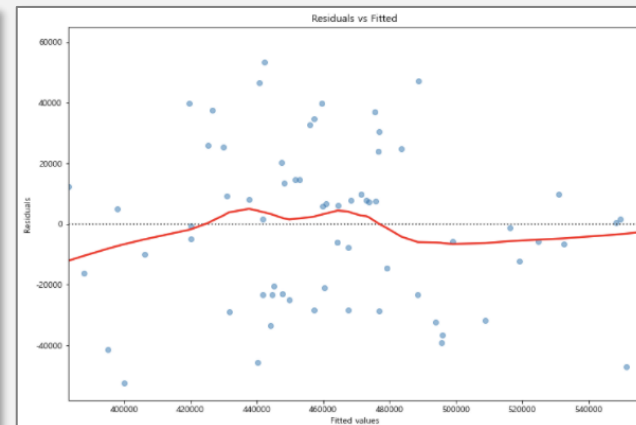
02 선형회귀 모델 구축 (OLS)

- 학습용 데이터셋과 테스트용 데이터셋 분리 (train_test_split)

(1) 선형회귀모델 가정확인



< 정규성 검정 >



< 등분산성 검정 >

- 정규성, 등분산성 모두 만족함을 확인할 수 있음 → **OLS 사용가능**

Part 2-2

다중회귀분석 : 자동차부품 수입액

(2) OLS Regression 결과 확인

OLS Regression Results

Dep. Variable:	수입액	R-squared:	0.691
Model:	OLS	Adj. R-squared:	0.654
Method:	Least Squares	F-statistic:	18.81
Date:	Thu, 31 Mar 2022	Prob (F-statistic):	6.31e-13
Time:	12:41:48	Log-Likelihood:	-777.51
No. Observations:	67	AIC:	1571.
Df Residuals:	59	BIC:	1589.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-2.013e+05	4.17e+05	-0.482	0.631	-1.04e+06	6.34e+05
환율	-257.0979	104.516	-2.460	0.017	-466.234	-47.962
소비자 물가지수	4066.9419	3513.570	1.157	0.252	-2963.695	1.11e+04
기준금리(연%)	-6.57e+04	1.58e+04	-4.161	0.000	-9.73e+04	-3.41e+04
국제유가	84.6599	584.301	0.145	0.885	-1084.524	1253.844
SCFI	0.5758	5.500	0.105	0.917	-10.431	11.582
고용률(%)	6138.1841	4212.830	1.457	0.150	-2291.670	1.46e+04
가동률지수(원지수)	2897.3764	406.989	7.119	0.000	2082.993	3711.760

- 유의수준 0.05를 넘는 p-value를 가진 피처변수 제거 필요
- 제거할 피처: 국제유가, SCFI, 고용률(%)

(3) 유의성 검정 후 피처변수 제거 / 모델 선언 및 학습

OLS Regression Results

Dep. Variable:	수입액	R-squared:	0.675
Model:	OLS	Adj. R-squared:	0.654
Method:	Least Squares	F-statistic:	32.25
Date:	Thu, 31 Mar 2022	Prob (F-statistic):	1.56e-14
Time:	12:42:01	Log-Likelihood:	-779.12
No. Observations:	67	AIC:	1568.
Df Residuals:	62	BIC:	1579.
Df Model:	4		
Covariance Type:	nonrobust		

가능한

3개의 변수

	coef	std err	t	P> t	[0.025	0.975]
const	5.585e+04	2.16e+05	0.259	0.796	-3.75e+05	4.87e+05
환율	-256.5119	82.376	-3.114	0.003	-421.180	-91.844
소비자 물가지수	5143.3360	1819.768	2.826	0.006	1505.671	8781.001
기준금리(연%)	-6.193e+04	1e+04	-6.183	0.000	-8.19e+04	-4.19e+04
가동률지수(원지수)	2985.0324	385.113	7.751	0.000	2215.203	3754.861

회귀계수 확인가능

3개의 피처 제거 후 나머지 변수에 대해 유의함을 확인

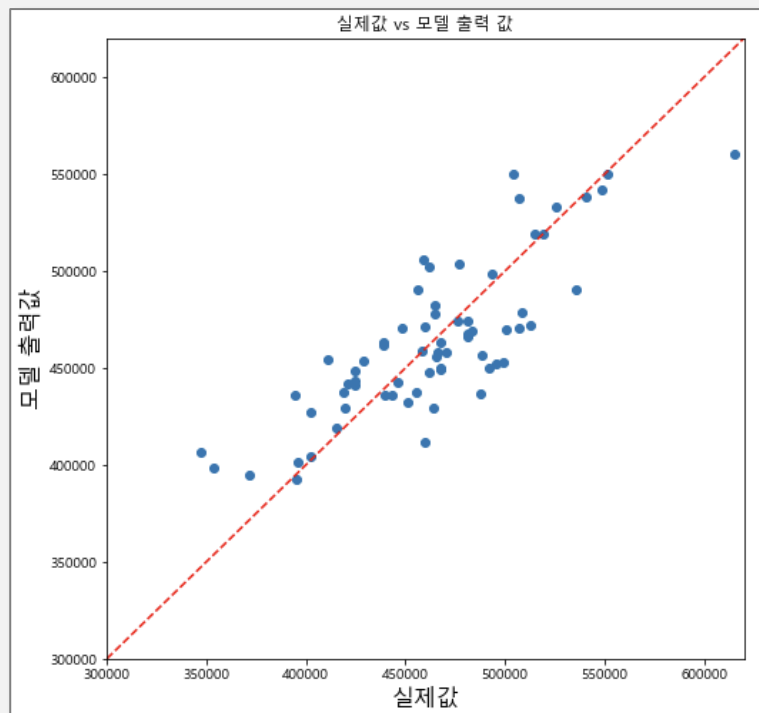
➤ 회귀식 도출

$$\begin{aligned} \text{수입액} = & -256.5119 * \text{환율} + 5143.3360 * \text{소비자 물가지수} \\ & - 6.193e+04 * \text{기준금리(연\%)} + 2985.0324 * \text{가동률지수(원지수)} \end{aligned}$$

다중회귀분석 : 자동차부품 수입액

03 예측 : 미래의 반응변수 값 예측하기

- 실제값과 모델 출력값이 거의 비슷함을 예측할 수 있음



04 해당 모델의 성능 평가

- MSE, RMSE, MAE, R2_Score 값 확인

테스트용 데이터셋 결과
MSE: 487779727.809
RMSE: 22085.736
MAE: 17309.141
R2_Score: 0.741

- ▶ 테스트용 데이터셋 성능평가 결과, 결정계수가 74.1%의 정확도
- ▶ 수출보다는 예측력이 조금 떨어지지만, 어느 정도 정확도는 있음
- ▶ 각각 RMSE가 22,085이므로, 예측값과 실제값의 차이가 이 수치만큼 차이날 수 있음을 예상할 수 있음

다중회귀분석 : 자동차부품 수입액

05 예측치와 실제 데이터 비교

- 회귀분석에 사용되지 않은 2022년 1월, 2월 데이터셋을 활용하여 실제값과 비교

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	항공운임	가동률지수(원지수)	고용률
0	2022.01	10816657	5739959	1205.50	104.69	1.25	2.79	122.9	59.6
1	2022.02	10381788	5493588	1197.83	105.30	1.25	2.85	122.6	60.6

- 위에서 산출한 회귀식에 해당 독립변수의 값들을 넣어 값을 예측

$$\begin{aligned} \text{수입액} = & -256.5119 * \text{환율} + 5143.3360 * \text{소비자 물가지수} \\ & - 6.193e+04 * \text{기준금리(연\%)} + 2985.0324 * \text{가동률지수(원지수)} \end{aligned}$$

- 예측 결과

예측 2022년 1월 수입액: 422859
실제 2022년 1월 수입액: 488239

예측 2022년 2월 수입액: 418113
실제 2022년 2월 수입액: 484919



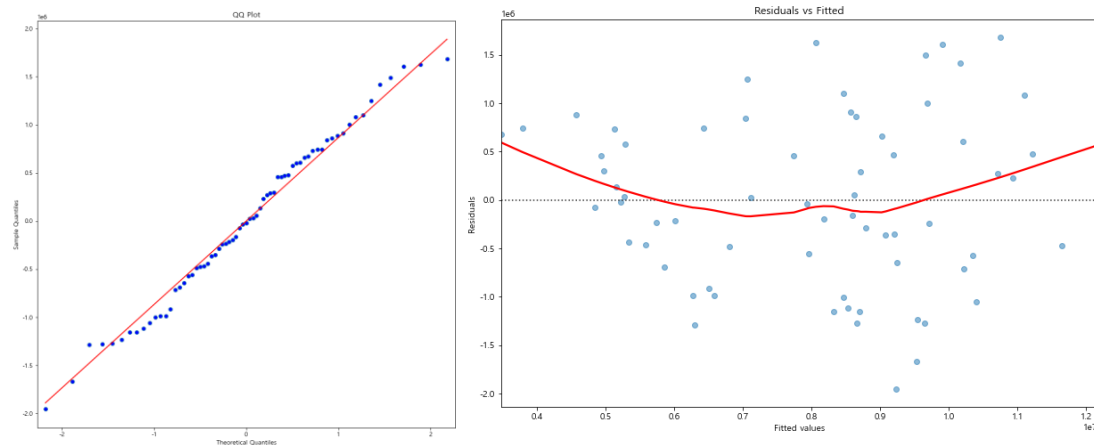
- 2022년 1~2월 자동차부품 수입액은 예측값과 실제값의 차이가 RMSE 값인 22,085보다는 더 큰 차이가 남
- 수출보다는 예측력이 조금 떨어짐을 확인할 수 있음

다중회귀분석 : 반도체 수출액

01 선형회귀 모델 구축 (OLS)

- 독립변수 X : 환율, 소비자 물가지수, 기준금리(연%), 국제유가, 항공운임, 고용률(%), 가동률지수(원지수)

(1) 선형회귀모델 가정확인



< 확률 오차의 정규성 >

< 확률 오차의 등분산성 >

- 정규성, 등분산성 모두 만족함을 확인할 수 있음 → **OLS 사용가능**

(2) 학습용 데이터셋 / 테스트용 데이터셋 분리

(3) 유의성 검정 후 피처변수 제거 / 모델 선언 및 학습

OLS Regression Results

Dep. Variable:	수출액	R-squared:	0.847
Model:	OLS	Adj. R-squared:	0.837
Method:	Least Squares	F-statistic:	85.60
Date:	Fri, 01 Apr 2022	Prob (F-statistic):	1.54e-24
Time:	01:44:56	Log-Likelihood:	-1012.6
No. Observations:	67	AIC:	2035.
Df Residuals:	62	BIC:	2046.
	4		
	nonrobust		

회귀계수 확인가능

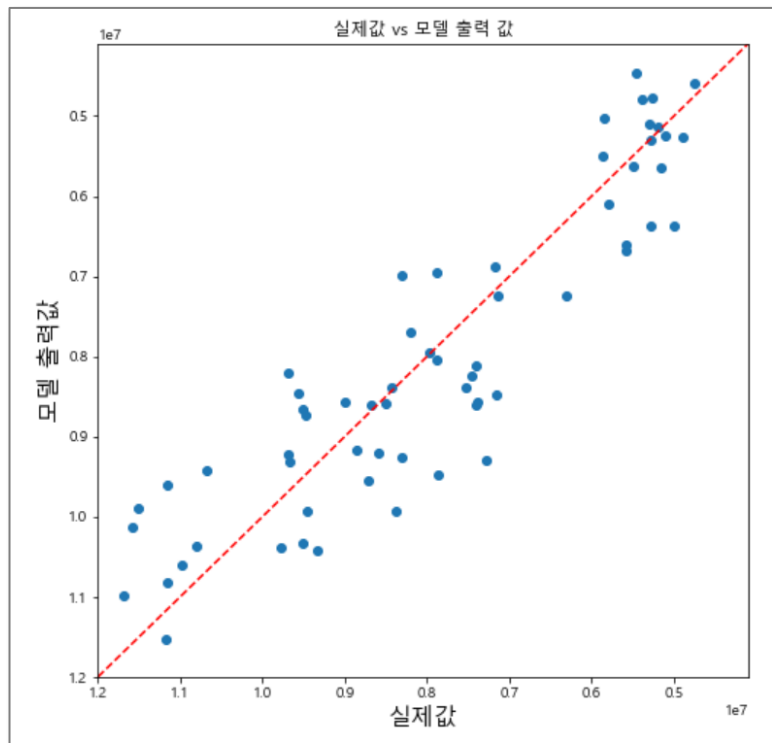
3개의 피처 제거 후 나머지 변수에 대해 유의함을 확인

	coef	std err	t	P> t	[0.025	0.975]
const	-6.078e+07	7.07e+06	-8.599	0.000	-7.49e+07	-4.67e+07
환율	-1.918e+04	2786.278	-6.885	0.000	-2.48e+04	-1.36e+04
소비자 물가지수	8.967e+05	7.15e+04	12.535	0.000	7.54e+05	1.04e+06
항공 운임	-7.75e+05	1.85e+05	-4.180	0.000	-1.15e+06	-4.04e+05
가동률지수(원지수)	3.274e+04	1.11e+04	2.951	0.004	1.06e+04	5.49e+04

다중회귀분석 : 반도체 수출액

02 예측 : 미래의 반응변수 값 예측하기

- 실제값과 모델 출력값이 거의 비슷함을 예측할 수 있음



03 해당 모델의 성능 평가

- MSE, RMSE, MAE, R2_Score 값 확인

테스트용 데이터셋 결과
MSE: 908456677417.802
RMSE: 953129.937
MAE: 850090.083
R2_Score: 0.798

- ▶ 테스트용 데이터셋 성능평가 결과, 결정계수가 79.8%의 정확도
- ▶ 각각 **RMSE가 953,129**이므로, **예측값과 실제값의 차이가 이 수치만큼 차이날 수 있음**을 예상할 수 있음

다중회귀분석 : 반도체 수출액

04 예측치와 실제 데이터 비교

- 회귀분석에 사용되지 않은 2022년 1월, 2월 데이터셋을 활용하여 실제값과 비교

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	항공운임	가동률지수(원지수)	고용률
0	2022.01	10816657	5739959	1205.50	104.69	1.25	2.79	122.9	59.6
1	2022.02	10381788	5493588	1197.83	105.30	1.25	2.85	122.6	60.6

- 위에서 산출한 회귀식에 해당 독립변수의 값들을 넣어 값을 예측

$$\text{수출액} = -1.918\text{e}+04 * \text{환율} + 8.967\text{e}+05 * \text{소비자물가지수} - 7.75\text{e}+05 * \text{항공운임} + 3.274\text{e}+04 * \text{가동률지수(원지수)} - 6.078\text{e}+07$$

- 예측 결과

예측 2022년 1월 수출액: 11835529
실제 2022년 1월 수출액: 10816657

예측 2022년 2월 수출액: 12473304
실제 2022년 2월 수출액: 10381788



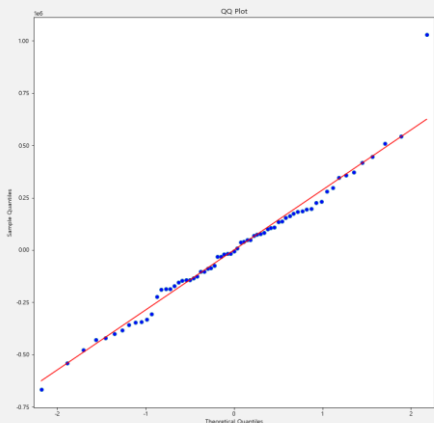
- 2022년 1~2월 반도체 수출액은 예측값과 실제값의 차이가 RMSE 값인 953,129와 유사하게 차이가 남
- 수출액을 거의 유사하게 예측함을 확인할 수 있음

다중회귀분석 : 반도체 수입액

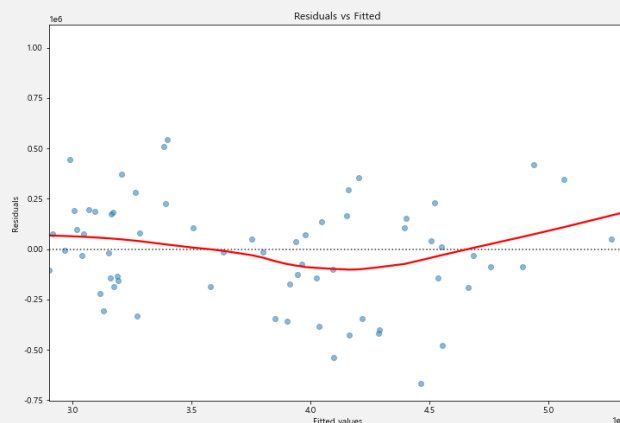
01 선형회귀 모델 구축 (OLS)

- 독립변수 X : 환율, 소비자 물가지수, 기준금리(연%), 국제유가, 항공운임, 고용률(%), 가동률지수(원지수)

(1) 선형회귀모델 가정확인



< 확률 오차의 정규성 >



< 확률 오차의 등분산성 >

- 정규성, 등분산성 모두 만족함을 확인할 수 있음 → **OLS 사용가능**

(2) 학습용 데이터셋 / 테스트용 데이터셋 분리

(3) 유의성 검정 후 피처변수 제거 / 모델 선언 및 학습

OLS Regression Results							
Dep. Variable:	수입액	R-squared:	0.843				
Model:	OLS	Adj. R-squared:	0.833				
Method:	Least Squares	F-statistic:	83.32				
Date:	Thu, 31 Mar 2022	Prob (F-statistic):	3.12e-24				
Time:	14:26:43	Log-Likelihood:	-937.89				
No. Observations:	67	AIC:	1886.				
Df Residuals:	62	BIC:	1886.				
	4						
	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	-1.576e+07	2.11e+06	-7.473	0.000	-2e+07	-1.15e+07	
소비자 물가지수	1.672e+05	2.41e+04	6.929	0.000	1.19e+05	2.15e+05	
기준금리(연%)	7.697e+05	1.94e+05	3.968	0.000	3.82e+05	1.16e+06	
항공 운임	5.568e+05	1.07e+05	5.206	0.000	3.43e+05	7.71e+05	
가동률지수(원지수)	1.126e+04	3936.786	2.861	0.006	3395.018	1.91e+04	

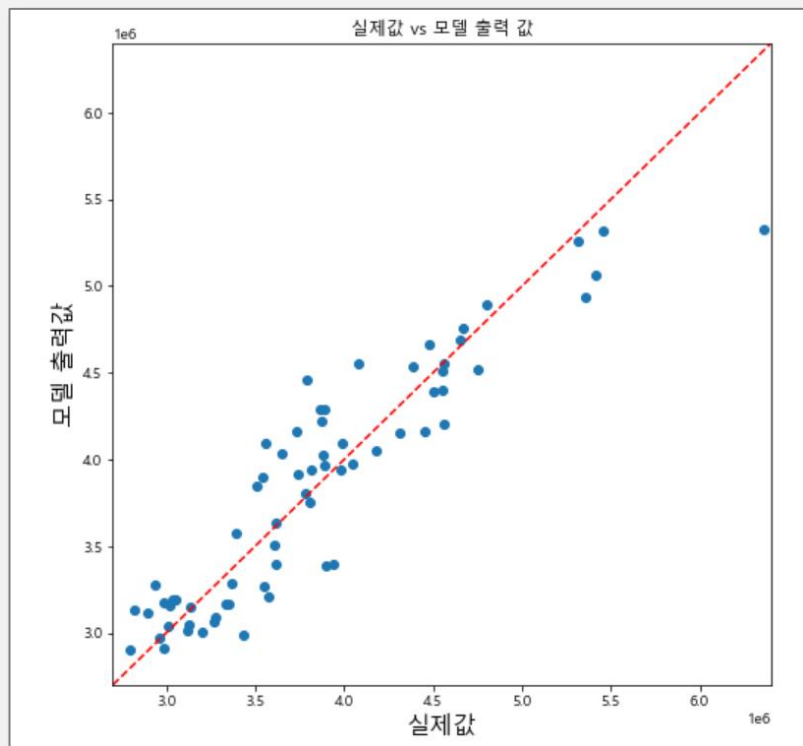
회귀계수 확인가능

3개의 피처 제거 후 나머지 변수에 대해 유의함을 확인

다중회귀분석 : 반도체 수입액

02 예측 : 미래의 반응변수 값 예측하기

- 실제값과 모델 출력값이 거의 비슷함을 예측할 수 있음



03 해당 모델의 성능 평가

- MSE, RMSE, MAE, R2_Score 값 확인

테스트용 데이터셋 결과
MSE: 135476203798.491
RMSE: 368070.922
MAE: 306412.937
R2_Score: 0.770

- ▶ 테스트용 데이터셋 성능평가 결과, 결정계수가 77.0%의 정확도
- ▶ 각각 **RMSE가 368,070**이므로, **예측값과 실제값의 차이가 이 수치만큼 차이날 수 있음**을 예상할 수 있음

다중회귀분석 : 반도체 수입액

04 예측치와 실제 데이터 비교

- 회귀분석에 사용되지 않은 2022년 1월, 2월 데이터셋을 활용하여 실제값과 비교

	시점	수출액	수입액	환율	소비자 물가지수	기준금리(연%)	항공운임	가동률지수(원지수)	고용률
0	2022.01	10816657	5739959	1205.50	104.69	1.25	2.79	122.9	59.6
1	2022.02	10381788	5493588	1197.83	105.30	1.25	2.85	122.6	60.6

- 위에서 산출한 회귀식에 해당 독립변수의 값들을 넣어 값을 예측

$$\begin{aligned} \text{수입액} = & 1.672\text{e}+05 * \text{소비자 물가지수} + 7.697\text{e}+05 * \text{기준금리(연\%)} \\ & + 5.568\text{e}+05 * \text{항공 운임} + 1.126\text{e}+04 * \text{가동률지수(원지수)} - 1.576\text{e}+07 \end{aligned}$$

- 예측 결과

예측 2022년 1월 수입액 : 5643619
실제 2022년 1월 수입액 : 5739959

예측 2022년 2월 수입액 : 5775641
실제 2022년 2월 수입액 : 5493588



- 2022년 1~2월 반도체 수입액은 예측값과 실제값의 차이가 RMSE 값인 368,070와 유사하게 차이가 남
- 수입액을 거의 유사하게 예측함을 확인할 수 있음

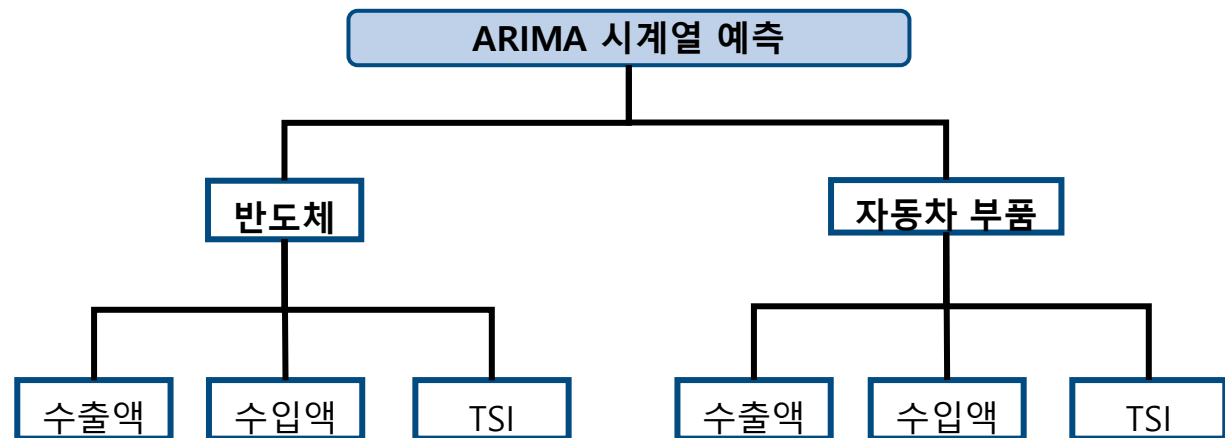
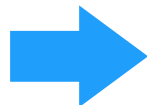
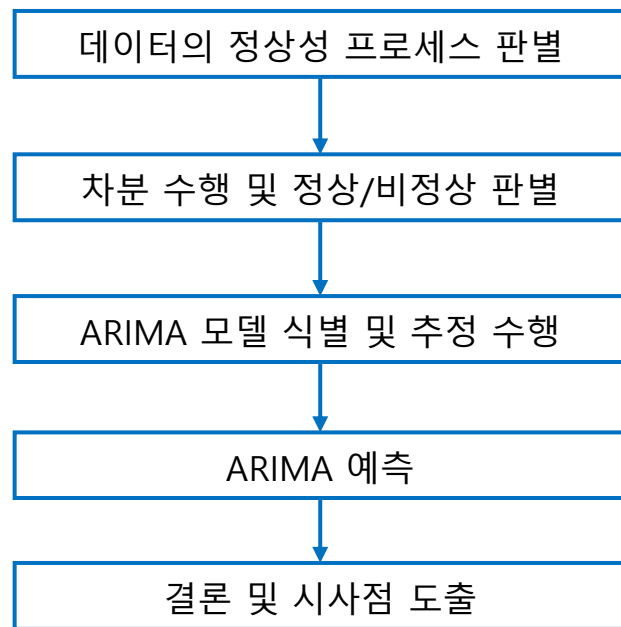
Part 2-3

ARIMA 모델링

- 자동차 부품
- 반도체
- TSI (자동차 부품, 반도체)
- RCA (자동차 부품, 반도체)

ARIMA 모형을 이용한 시계열 분석 개념도

★ ARIMA 분석 프로세스 개념도

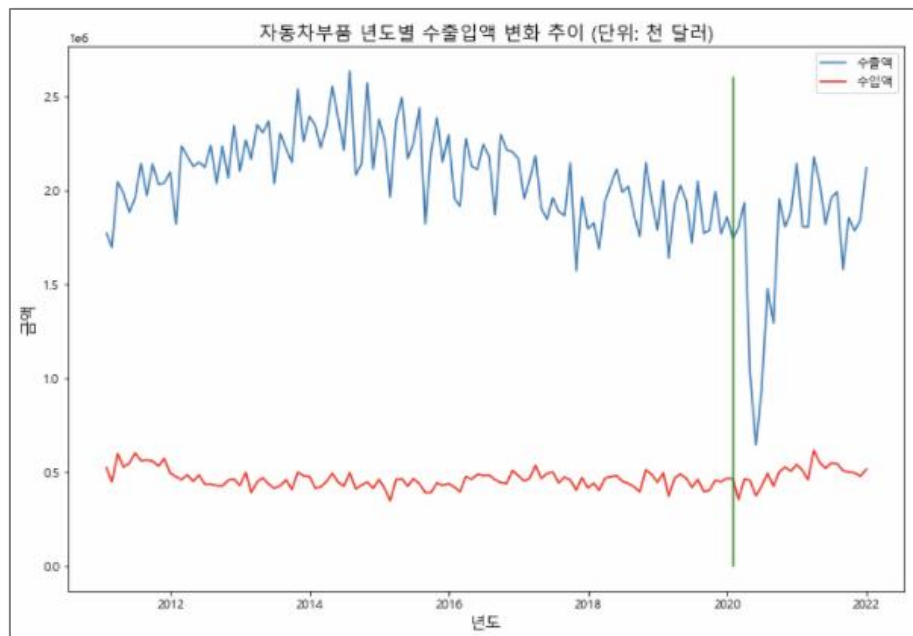


- 반도체와 자동차 부품의 수출액, 수입액, TSI를 **ARIMA모형**을 활용하여 시계열 분석 및 예측 진행

ARIMA : 자동차부품 산업 수출액 모델링

01 자동차부품 수출입액 변화 추이

- 대체로 수출액은 20억 달러를 상회하고 있으나, 코로나 19가 시작된 2020년 초반 수출액이 급격하게 감소하고 서서히 회복되고 있음



02 데이터의 정상성 확인 및 차분

원 데이터 ADF Statistic: -1.163	
원 데이터 p-value: 0.689	

1차 차분 ADF Statistic: -4.645	
1차 차분 p-value: 0.000	

- 시계열 분석을 하기 위해서는 '정상성' 을 만족해야 함
- 일반적으로 ADF Test 결과 p-value가 0.05보다 작으면 정상성이 있음
- 이에 따라, 1차 차분한 결과가 정상성이 있음 (차분모델은 정상성 만족)

03 p, d, q 파라미터 추정

- (p, d, q)가 (2, 1, 0)일 때, AIC의 값이 가장 작으므로 해당 파라미터 값을 활용하여 **ARIMA (2, 1, 0)**을 활용

ARIMA : 자동차부품 산업 수출액 모델링

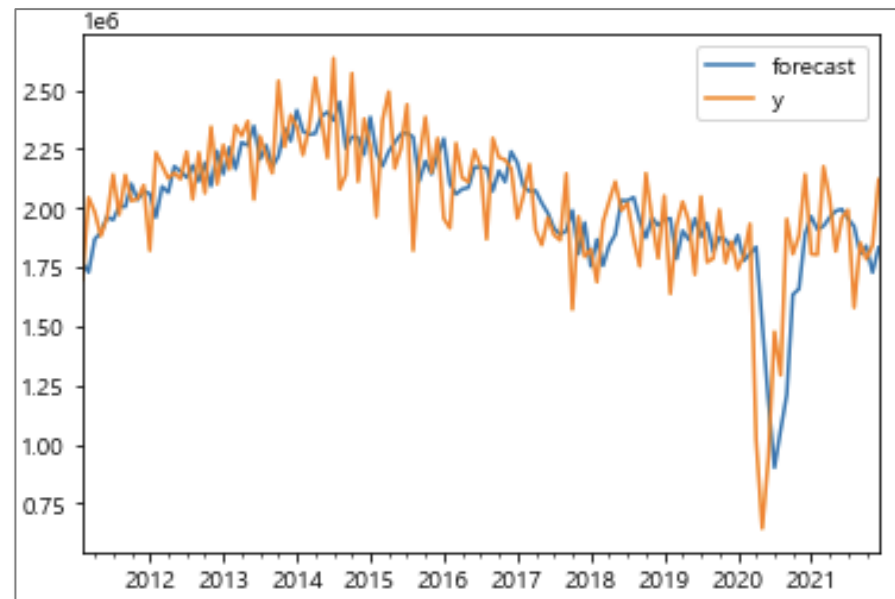
04 p-value 및 AIC값 확인

- ARIMA(2, 1, 0) 모델 수행 결과, AIC는 3594.722이고, **모델은 유의함**

ARIMA Model Results						
Dep. Variable:	D.y	No. Observations:	131			
Model:	ARIMA(2, 1, 0)	Log Likelihood	-1794.361			
Method:	css-mle	S.D. of innovations	214630.872			
Date:	Thu, 31 Mar 2022	AIC	3594.722			
Time:	22:12:40	BIC	3603.347			
Sample:	02-28-2011	HQIC	3598.226			
	- 12-31-2021					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	-0.5815	0.082	-7.088	0.000	-0.742	-0.421
ar.L2.D.y	-0.3533	0.082	-4.303	0.000	-0.514	-0.192
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-0.8230	-1.4674j	1.6824	-0.3313		
AR.2	-0.8230	+1.4674j	1.6824	0.3313		

05 ARIMA 모델 학습 후 예측된 값 그래프 시각화

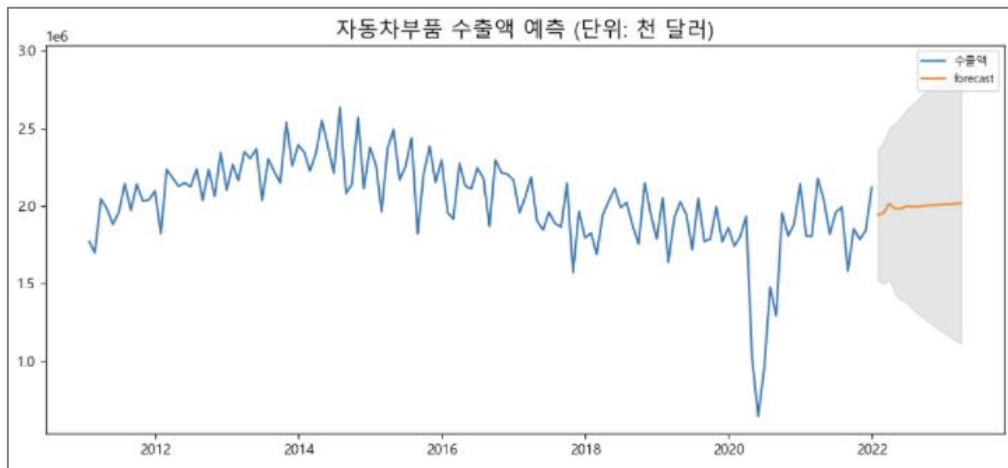
- y는 실제 수출액, forecast는 예측된 수출액 값을 의미
- 실제와 비교해 보았을 때, 큰 차이가 있지 않음



ARIMA : 자동차부품 산업 수출액 모델링

06 2022.01.01~2023.03.31 수출액 예측

- 2022년 이후로 15개월 동안의 수출액을 예측해보았을 때, 코로나 이후로 수출액이 크게 감소했지만, 다시 평균지점을 회복하는 수출액 동향을 보일 것으로 예측됨



ARIMA : 자동차부품 산업 수입액 모델링

01 데이터의 정상성 확인 및 차분

원 데이터 ADF Statistic: -3.155
원 데이터 p-value: 0.023

1차 차분 ADF Statistic: -3.535
1차 차분 p-value: 0.007

- ADF Test 결과 p-value가 0.05보다 작으므로, 원 데이터 모델과 1차 차분한 모델 모두 정상성의 가정을 만족

02 p, d, q 파라미터 추정

- (p, d, q)가 (2, 1, 0)일 때, AIC의 값이 가장 작으므로 해당 파라미터 값을 활용하여 **ARIMA (2, 1, 0)**을 활용

03 p-value 및 AIC값 확인

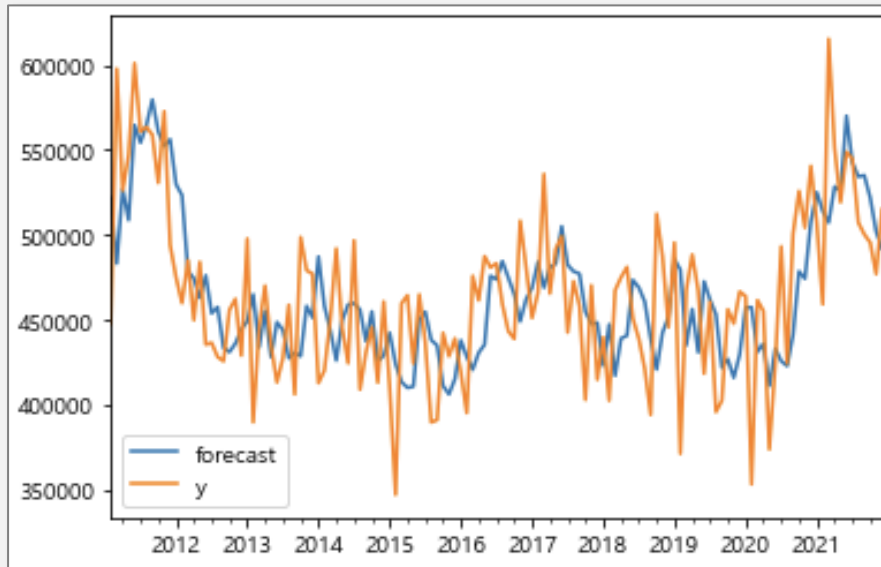
- ARIMA(2, 1, 0) 모델 수행 결과, AIC는 3154.223이고, **모델은 유의함**

ARIMA Model Results						
Dep. Variable:	D.y	No. Observations:	131			
Model:	ARIMA(2, 1, 0)	Log Likelihood	-1574.112			
Method:	css-mle	S.D. of innovations	39914.122			
Date:	Wed, 30 Mar 2022	AIC	3154.223			
Time:	23:57:34	BIC	3162.849			
Sample:	02-28-2011 - 12-31-2021	HQIC	3157.728			
	coef	std err	z	P> z	[0.025	0.975]
ar.L1.D.y	-0.6929	0.078	-8.840	0.000	-0.847	-0.539
ar.L2.D.y	-0.4470	0.081	-5.550	0.000	-0.605	-0.289
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-0.7750	-1.2792j	1.4956	-0.3367		
AR.2	-0.7750	+1.2792j	1.4956	0.3367		

ARIMA : 자동차부품 산업 수입액 모델링

04 ARIMA 모델 학습 후 예측된 값 그래프 시각화

- y는 실제 수입액, forecast는 예측된 수입액 값을 의미
- 실제와 비교해 보았을 때, 큰 차이가 있지 않음



05 2022.01.01~2023.03.31 수입액 예측

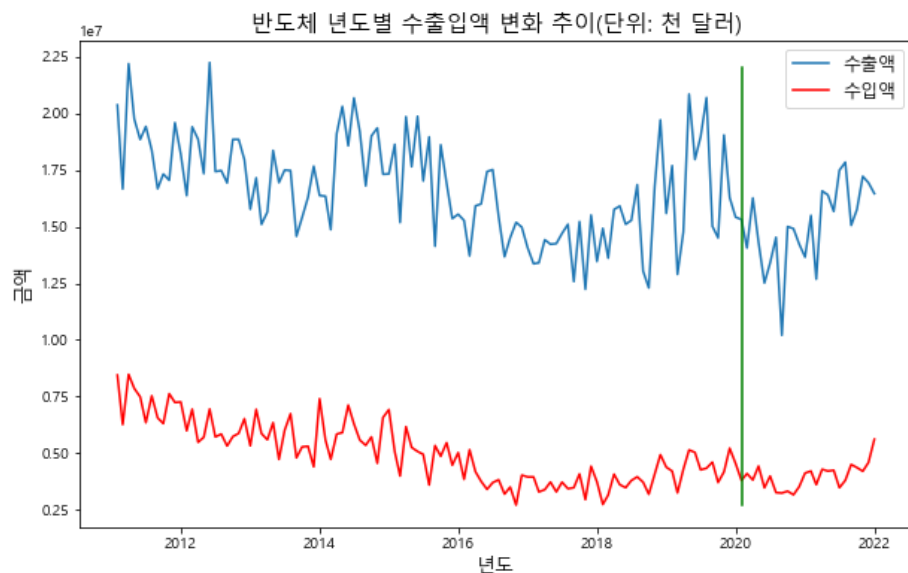
- 2022년 이후로 15개월 동안의 수입액을 예측해보았을 때, 수입액은 코로나 19에 영향을 크게 받지 않지만, 앞으로의 수입액은 전보다는 조금 안정된 모형으로 수렴할 것으로 예측됨



ARIMA : 반도체 산업 수출액 모델링

01 반도체 수출입액 변화 추이

- 코로나 19가 시작된 2020년 초반 수출액이 급격하게 감소함



02 데이터의 정상성 확인 및 차분

원 데이터 ADF Statistic: -2.363
원 데이터 p-value: 0.153
1차 차분 ADF Statistic: -6.109
1차 차분 p-value: 0.000

- 시계열 분석을 하기 전, '정상성', '정규성' 만족을 위해 차분
- 일반적으로 ADF Test 결과 p-value가 0.05보다 작으면 정상성이 있음
- 이에 따라, 1차 차분한 결과가 정상성이 있음 (차분모델은 정상성 만족)

03 p, d, q 파라미터 추정

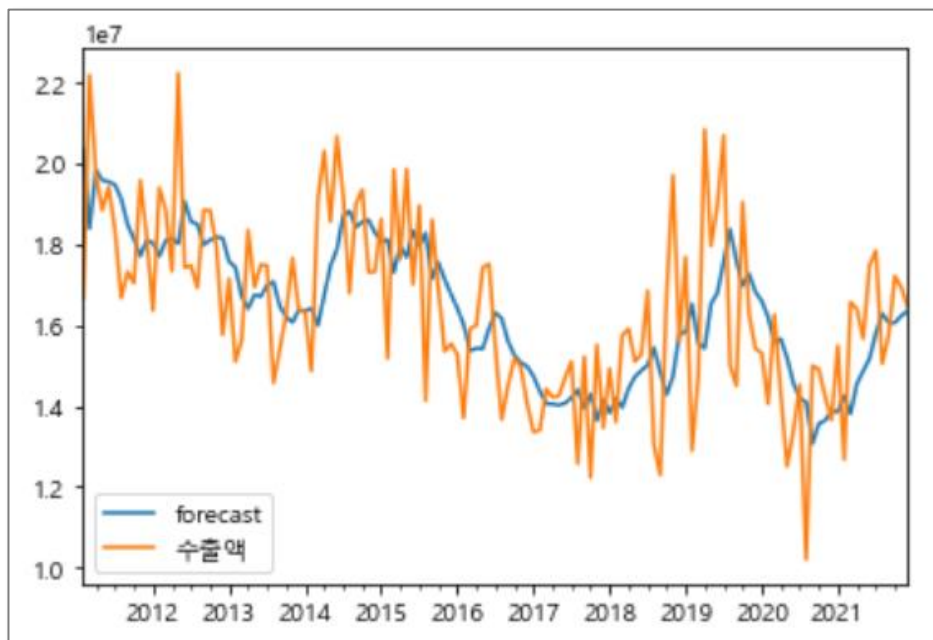
- auto_arima를 사용하여 ARIMA (2, 1, 1)를 구함

```
Best model: ARIMA(2,1,1)(0,0,0)[0] intercept  
Total fit time: 2.178 seconds  
ARIMA(order=(2, 1, 1), scoring_args={}, suppress_warnings=True)
```

ARIMA : 반도체 산업 수출액 모델링

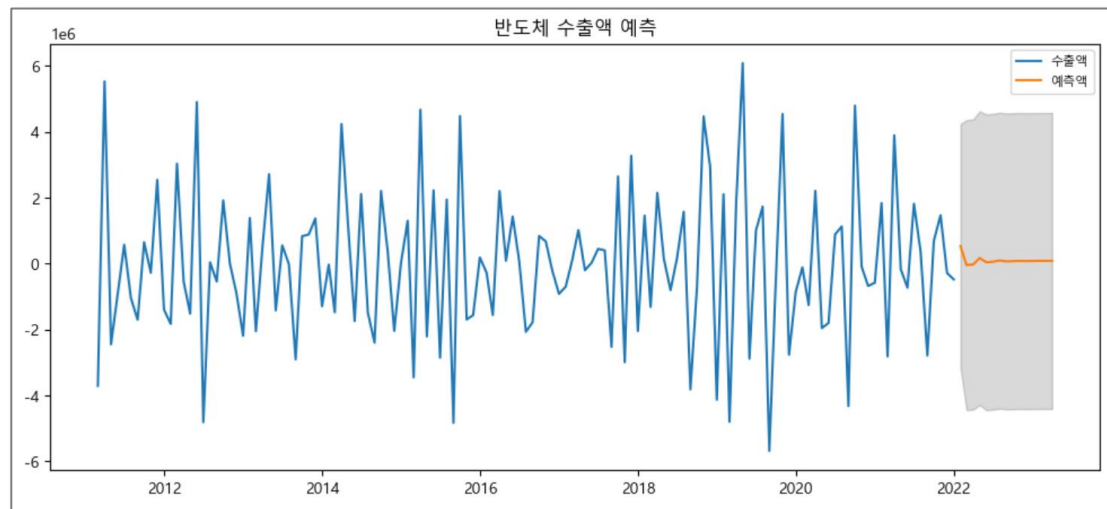
04 ARIMA 모델 학습 후 예측된 값 그래프 시각화

- ARIMA(2, 1, 0) 모델 수행 결과, **모델은 유의함**
- 실제값과 예측값을 비교했을 때, 거의 비슷한 흐름을 나타냄



05 2022.01.01~2023.03.31 수출액 예측

- 2022년 이후로 15개월 동안의 수출액을 예측해보았을 때, 코로나 전후로 반도체 수출액이 큰 폭으로 변화하지만, 앞으로 평균지점을 회복하는 수출액 동향을 보일 것으로 예측됨



ARIMA : 반도체 산업 수입액 모델링

01 데이터의 정상성 확인 및 차분

```
원 데이터 ADF Statistic: -2.353
원 데이터 p-value: 0.155
1차 차분 ADF Statistic: -3.849
1차 차분 p-value: 0.002
```

- ADF Test 결과 p-value가 0.05보다 작으므로, 1차 차분한 모델은 정상성의 가정을 만족한다.

02 p, d, q 파라미터 추정

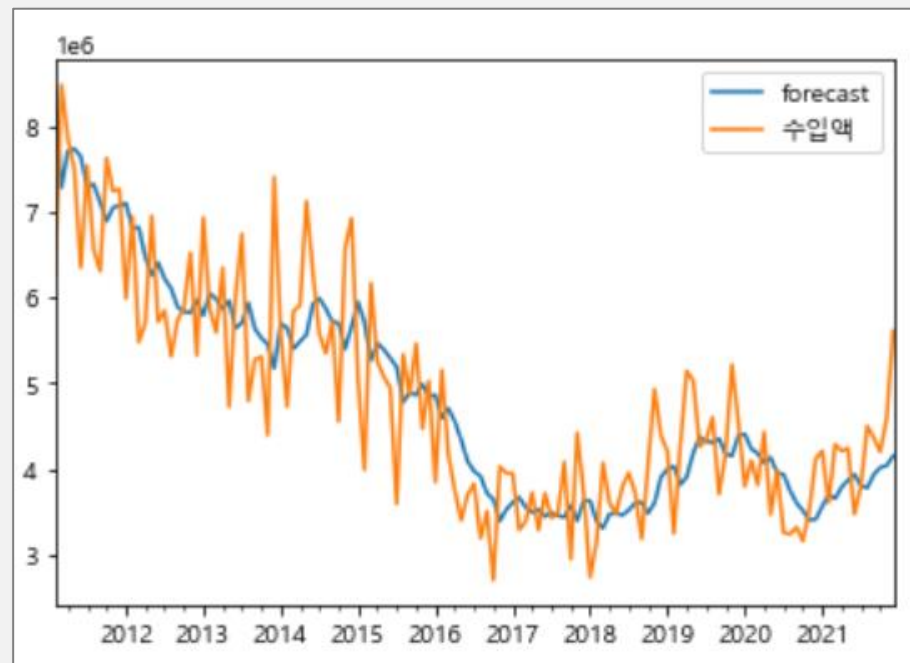
- auto_arima를 사용하여 ARIMA (0, 1, 1)를 구함

```
Best model: ARIMA(0,1,1)(0,0,0)[0] intercept
Total fit time: 2.489 seconds

ARIMA(order=(0, 1, 1), scoring_args={}, suppress_warnings=True)
```

03 ARIMA 모델 학습 후 예측된 값 그래프 시각화

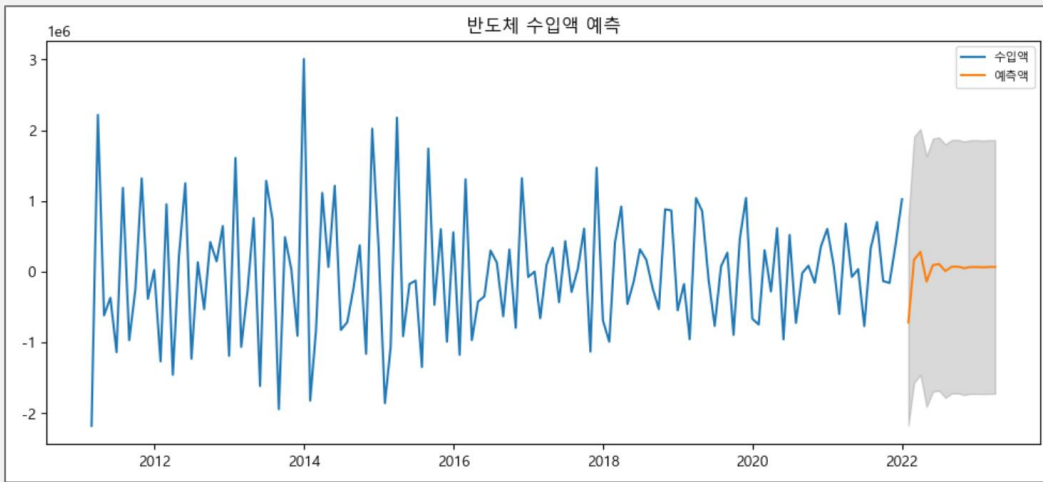
- ARIMA(0, 1, 1) 모델 수행 결과, **모델은 유의함**
- 실제값과 예측값을 비교했을 때, 실제 값보다는 조금 낮게 형성



ARIMA : 반도체 산업 수입액 모델링

04 2022.01.01~2023.03.31 수입액 예측

- 2022년 이후로 15개월 동안의 수입액을 예측해보았을 때, 코로나 전에는 반도체 수입액이 큰 폭으로 변화했지만, 앞으로 평균지점을 회복하는 수입액 동향을 보일 것으로 예측됨



RCA 지수 분석 : 자동차부품 품목

01 자동차 부품 산업 품목 선정 기준

- 선행연구 논문을 참고하여 국제 무역 코드 HSCODE 4단위 기준 '차량용 부분품, 부속품'으로 선택 → HSCODE 6단위로 조사
- 선행연구 논문 참고 라공우, 송진구.(2017), 미국시장에 대한 한·중 자동차부품산업의 국제경쟁력에 관한 연구

02 비교 대상 국가 선정

- 비교 대상 국가 : 독일, 미국, 중국
- 대표적인 자동차 산업 강국인 독일과, 세계 무역의 큰 부분을 차지하는 미국과 중국을 비교대상 국가로 선정하였음.



RCA 지수 분석 : 자동차부품 품목

01 2017 - 2018년도 RCA 분석

년도	품목명	한국	독일	미국	중국
2017	완충기	0.25	2.89	1.18	0.41
	안전벨트	1.21	1.93	0.54	0.20
	기타	1.15	2.33	1.65	0.55
	기어박스	1.60	3.18	1.36	0.20
	차동장치 및 드라이브축	0.93	2.54	1.20	0.31
	로드휠	0.88	1.75	0.82	2.41
	서스펜션	0.83	2.59	1.48	1.07
	라디에이터	0.40	1.60	0.79	1.54
	소음기와 배기기관	0.52	2.68	0.94	0.57
	클러치	1.47	3.57	1.03	0.53
	운전대	1.50	2.18	1.35	0.70
	부분품	3.12	1.39	1.47	0.40

년도	품목명	한국	독일	미국	중국
2018	완충기	0.23	2.93	1.19	0.36
	안전벨트	0.92	1.92	0.50	0.19
	기타	1.04	2.33	1.45	0.52
	기어박스	1.43	3.15	1.28	0.19
	차동장치 및 드라이브축	0.92	2.52	1.16	0.30
	로드휠	0.72	1.70	0.68	2.27
	서스펜션	0.74	2.55	1.28	1.05
	라디에이터	0.38	1.66	0.69	1.42
	소음기와 배기기관	0.45	2.59	0.80	0.56
	클러치	1.12	3.39	1.00	0.48
	운전대	1.41	2.05	1.22	0.67
	부분품	2.74	1.35	1.17	0.39

- 2017년과 2018년도 한국은 부분품이 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2018년에 중국, 독일, 미국은 각각 클러치, 기타, 로드휠이 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2018년에 한국의 안전벨트, 클러치, 부분품의 RCA값이 0.3씩 줄어듦

RCA 지수 분석 : 자동차부품 품목

02 2019 - 2020년도 RCA 분석

년도	품목명	한국	독일	미국	중국
2019	완충기	0.25	2.71	1.19	0.37
	안전벨트	1.29	1.83	0.40	0.19
	기타	1.16	2.15	1.44	0.56
	기어박스	1.49	3.20	1.29	0.17
	차동장치 및 드라이브축	1.14	2.49	1.22	0.30
	로드휠	0.86	1.77	0.68	2.09
	서스펜션	0.80	2.45	1.33	1.08
	라디에이터	0.33	1.69	0.70	1.45
	소음기와 배기기관	0.49	2.25	0.84	0.58
	클러치	1.37	3.16	1.05	0.50
	운전대	1.70	1.94	1.13	0.69
	부분품	3.32	1.24	1.16	0.43

년도	품목명	한국	독일	미국	중국
2020	완충기	0.29	2.92	1.16	0.43
	안전벨트	1.00	2.10	0.41	0.19
	기타	1.14	2.34	1.47	0.74
	기어박스	1.45	3.64	1.29	0.18
	차동장치 및 드라이브축	1.31	2.84	1.23	0.38
	로드휠	0.72	1.98	0.68	2.05
	서스펜션	0.87	2.71	1.39	1.13
	라디에이터	0.30	1.85	0.68	1.49
	소음기와 배기기관	0.37	2.33	0.74	0.62
	클러치	1.53	3.61	1.09	0.55
	운전대	1.78	2.17	1.19	0.72
	부분품	3.42	1.36	1.25	0.47

- 2018년과 2019년도 한국은 부분품이 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2019년 한국 부분품의 RCA값이 0.57 증가
- 2020년에 중국, 독일, 미국은 각각 기어박스, 기타, 로드휠이 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2020년 한국 차동장치 및 드라이브 축이 수출량이 증가하여 2017년의 수입경향에서 수출경향으로 전환됨

TSI 및 RCA 지수 분석 : 반도체 품목

01 반도체 산업 데이터 : MTI코드, HS코드 분류표

MTI			HS 6단위
반도체 (831)	집적회로 반도체	메모리반도체(831110)	847330 854232
		시스템반도체	
		프로세서와 콘트롤러(831120)	854231
		증폭기(831130)	854233
		기타 집적회로반도체(831190)	852352 854239
	집적회로반도체 부품	집적회로반도체 부품(831200)	854290 854390 854890
	개별소자 반도체	트랜지스터(831310)	854121 854129
		다이오드(831320)	854110
		기타 개별소자반도체(831390)	854130 854140 854150 854160
	개별소자 반도체 부품	개별소자반도체 부품(831400)	854190
	실리콘웨이퍼	실리콘웨이퍼(831500)	381800

출처 : 김은영.(2021),한국과 주요국 간의 반도체산업 수출경쟁력 및 수출경합도 비교 분석.

02 지수 분석 해당 국가 선정 이유

- 세계 주요 반도체 기업이 유치되어 있는 미국, 대만, 중국을 대상으로 적절 판단

세계 주요 반도체 기업 순위 (2020년)

순위	기업명	본사	유형
1	Intel	미국	IDM
2	Samsung	한국	IDM
3	TSMC	대만	파운드리

출처 : IC Insight와 각 기업 사이트 참고

RCA 지수 분석 : 반도체 품목

01 2017 - 2018년도 RCA 분석

년도	품목명	한국	중국	대만	미국
2017	메모리반도체	7.468672	1.70423	6.420565	1.319125
	시스템반도체	0.28333	2.080509	2.346053	0.482004
	집적회로반도체 부품	6.845199	0.734301	4.131291	0.848786
	트랜지스터	1.062239	1.479941	3.917929	0.804951
	다이오드	0.705898	1.35802	4.549857	0.934782
	기타 개발소자반도체	2.261636	2.185195	3.431621	0.705037
	개별소자반도체 부품	0.814591	1.067853	3.276051	0.673075
	실리콘웨이퍼	1.629974	2.238256	4.233066	0.869696

년도	품목명	한국	중국	대만	미국
2018	메모리반도체	8.473355	1.911852	5.4624	1.102312
	시스템반도체	0.292584	2.443634	2.137409	0.431329
	집적회로반도체 부품	5.324552	0.838881	3.672149	0.74104
	트랜지스터	1.130509	1.395075	4.00495	0.808199
	다이오드	0.823237	1.483235	5.009662	1.01095
	기타 개발소자반도체	2.233798	2.464785	3.759166	0.7586
	개별소자반도체 부품	0.559109	1.171988	2.720717	0.549041
	실리콘웨이퍼	2.763085	1.66418	4.879059	0.984594

- 2017년과 2018년도 한국은 메모리 반도체가 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2018년에 한국의 메모리 반도체의 RCA값이 전년대비 약 1.004 증가하였음
- 2018년에 중국, 대만, 미국은 각각 시스템반도체, 기타 개발소자반도체, 메모리반도체가 여타 품목보다 경쟁력이 있는 것으로 나타남

RCA 지수 분석 : 반도체 품목

02 2019 - 2020년도 RCA 분석

년도	품목명	한국	중국	대만	미국
2019	메모리반도체	7.19988	2.033687	6.955919	1.397535
	시스템반도체	0.34416	1.689522	1.720698	0.345711
	집적회로반도체 부품	0.561585	1.305595	5.333703	1.071611
	트랜지스터	1.317658	1.397023	3.567222	0.716702
	다이오드	0.894398	1.524503	4.564058	0.916979
	기타 개발소자반도체	1.909834	2.799814	3.16042	0.63497
	개별소자반도체 부품	0.561234	1.208967	2.088316	0.41957
	실리콘웨이퍼	2.39012	1.389218	5.135576	1.031804

년도	품목명	한국	중국	대만	미국
2020	메모리반도체	6.85272	1.787018	6.280132	1.523497
	시스템반도체	0.399351	1.34175	1.361481	0.330282
	집적회로반도체 부품	0.52182	1.229002	3.37485	0.818705
	트랜지스터	1.239281	1.417447	2.906995	0.705208
	다이오드	0.761592	1.454437	2.629715	0.637942
	기타 개발소자반도체	1.658356	2.420771	2.607858	0.63264
	개별소자반도체 부품	0.581338	0.921825	1.763382	0.427779
	실리콘웨이퍼	2.20498	1.280422	4.512591	1.094709

- 2019년과 2020년도 한국은 메모리 반도체가 여타 품목보다 경쟁력이 있는 것으로 나타남
- 2020년에 중국, 대만, 미국은 각각 기타 개발소자반도체, 메모리반도체가 여타 품목보다 경쟁력이 있는 것으로 나타남

ARIMA : TSI (자동차부품)

01 TSI 결과 및 예측분석

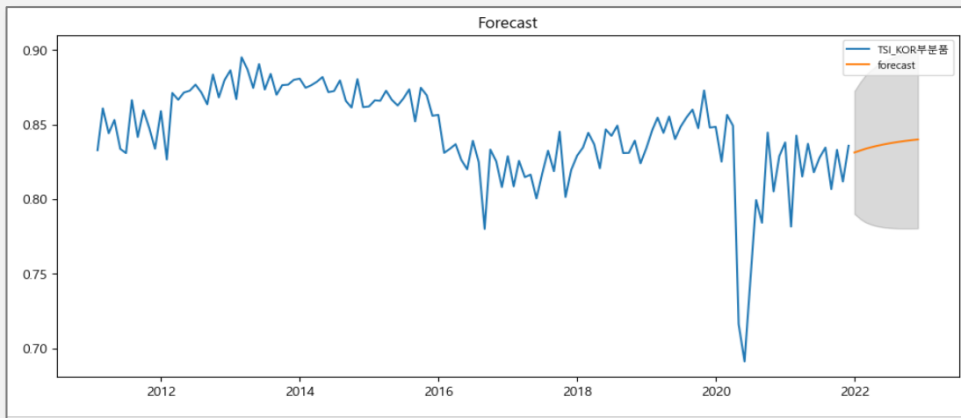
	한국	독일	미국	중국	최대 수출국	수입비중 높은 국가
완충기	- 0.465	0.339	- 0.319	- 0.027	독일	한국
안전벨트	0.121	- 0.146	- 0.649	- 0.452	한국	미국
기타	0.668	0.105	- 0.417	0.251	한국	미국
기어박스	0.439	0.494	- 0.182	- 0.664	독일	독일
차동장치 및 드라이브축	0.863	0.310	- 0.412	0.216	한국	미국
로드휠	0.006	0.045	- 0.617	0.953	중국	미국
서스펜션	0.570	0.237	- 0.440	0.460	한국	미국
라디에이터	0.364	- 0.085	- 0.443	0.709	중국	미국
소음기와 배기기관	0.395	0.337	- 0.483	0.425	중국	미국
클러치	0.663	0.408	- 0.224	0.176	한국	미국
운전대	0.541	0.208	- 0.406	0.226	한국	미국
부분품	0.840	0.175	- 0.306	0.338	한국	미국

- 차동장치 및 드라이브축(870850), 부분품(870899), 기타(870829), 클러치(870893), 서스펜션(870880), 운전대(870894) 품목이 수출특화에 가까운 품목이 될 것으로 예측됨
- 완충기(870810)는 수입특화에 가까운 품목으로 예측됨

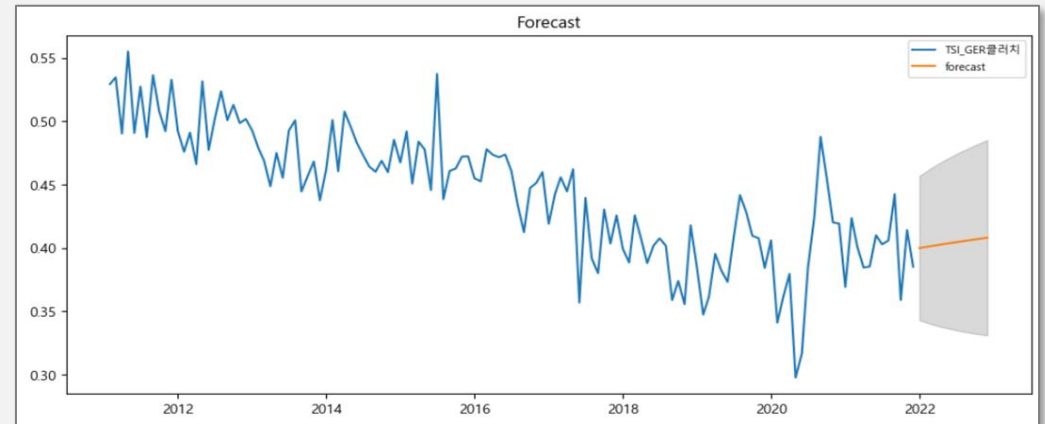
Part 2-3

ARIMA : TSI (자동차부품) 예측 분석

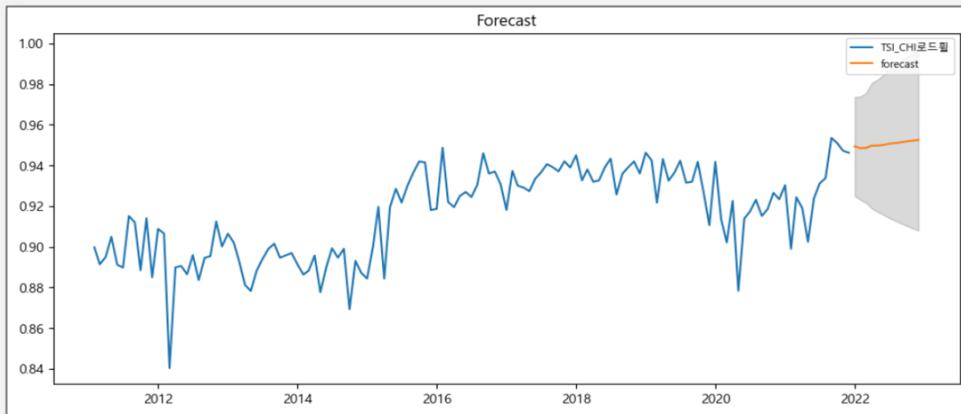
01 한국-부분품



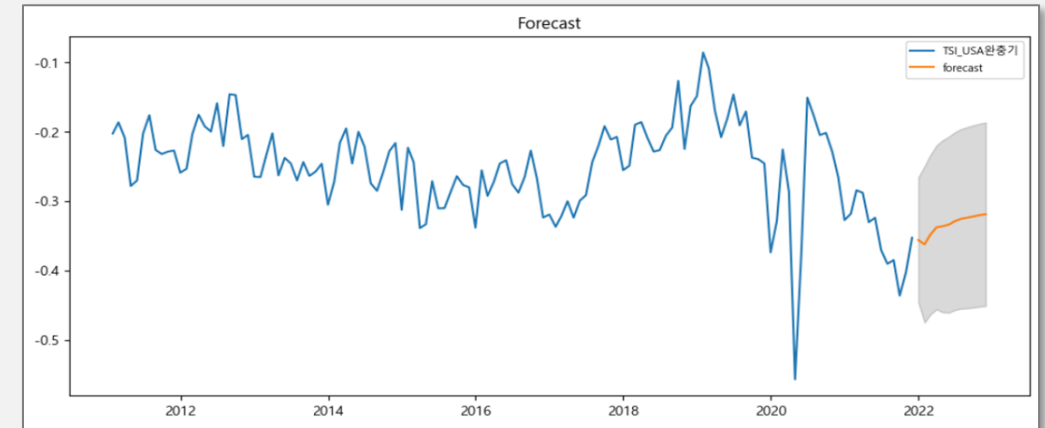
02 독일-클러치



03 중국-로드휠



04 미국-완충기



ARIMA : TSI (반도체)

01 TSI 결과 및 예측분석

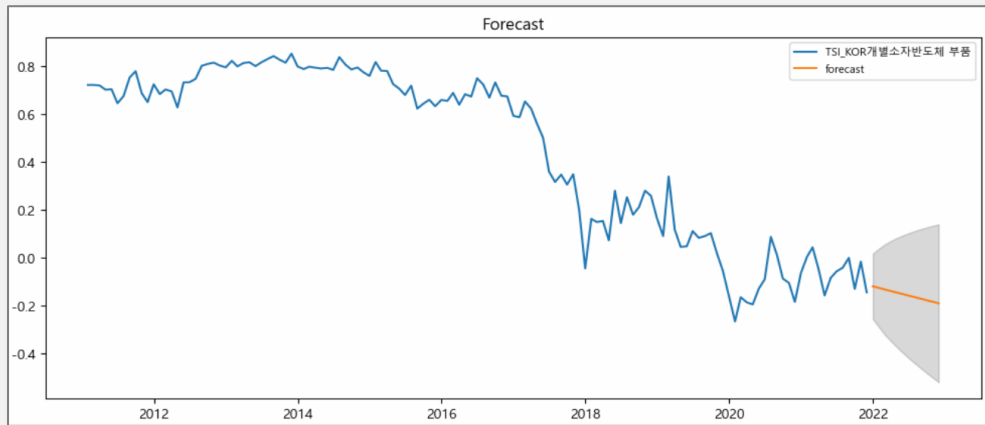
품목별	한국	중국	대만	미국	최대 수출국	수입비중 높은 국가
메모리반도체	0.526846	-0.1330245	-0.06111	-0.0872075	한국	중국
시스템반도체	0.048258	-0.5809818	0.406469	0.081809	대만	중국
집적회로반도체 부품	-0.05045	0.33738407	0.143401	0.040728	중국	한국
트랜지스터	0.037545	-0.16389	0.027941	0.093285	미국	중국
다이오드	-0.31913	-0.20442	0.395328	0.153038	대만	한국
기타 개발소자반도체	-0.20344	-0.52448	0.315954	-0.34133	대만	중국
개별소자반도체 부품	-0.19007	-0.43969	0.245037	0.086569	대만	중국
실리콘웨이퍼	-0.36702	0.122833	-0.41486	-0.12503	중국	대만

- 메모리 반도체 경우, 한국이 가장 높은 무역특화지수를 예측하고 있지만 **집적회로반도체 부품**과 **다이오드**는 해외 수입 의존도가 높아진 것으로 예측됨

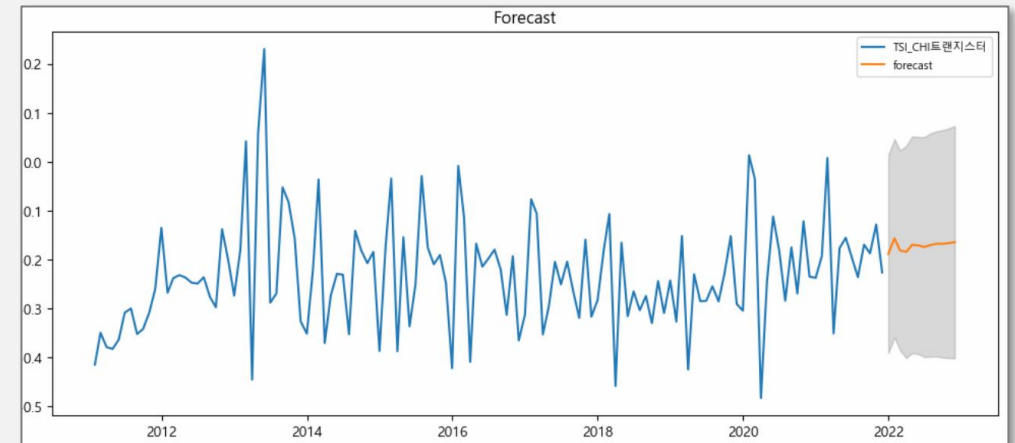
Part 2-3

ARIMA : TSI (반도체) 예측 분석

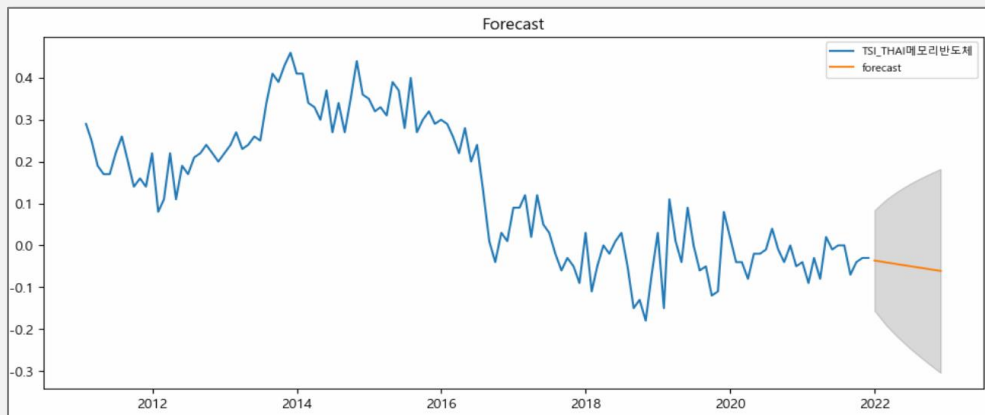
01 한국-개별소자반도체 부품



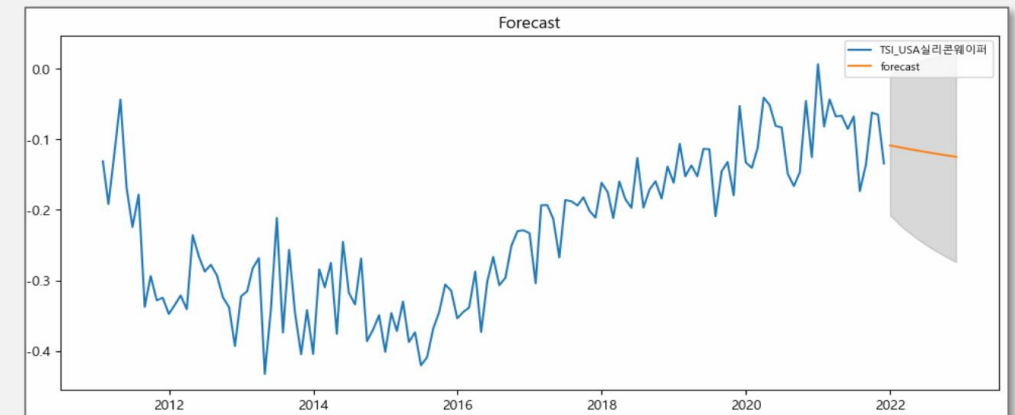
02 중국-트랜지스터



03 대만-메모리반도체



04 미국-실리콘웨이퍼



Part 3

결론 및 기대효과

- 수출입 금액 예측 및 경쟁력에 대한 최종적인 결론
- 개선사항 및 기대효과

수출입 금액 예측 및 경쟁력에 대한 최종적인 결론

수출입 금액 예측 모델링 구현에 대한 결론

- 다중회귀분석에서 **RMSE 값에 맞춰 수출입 금액의 결과가 예측이 잘 되었음**. 하지만, 수출입 금액에 영향을 주는 수많은 피처가 있으므로, **새로운 피처를 찾아 더 나은 예측력을 보일 필요**가 있음
- ARIMA에서는 2022~2023.3.31까지 예측을 했는데 코로나 19 장기화됨에 따라 수출입 금액의 **불확실성 증가된 가운데 정교한 학습과 데이터 확보가 필요함**

경쟁력 예측

- 자동차부품의 TSI 결과, **차동장치 및 드라이브축**, 부분품, 기타, 클러치, 서스펜션, 운전대 품목이 **수출특화에 가까운 품목**이 될 것으로 예측됨. 완충기는 수입특화에 가까운 품목으로 예측됨.
- 반도체의 TSI 결과, **메모리 반도체** 경우, 한국이 가장 높은 무역특화지수를 예측되고 있지만 **집적회로반도체 부품**과 **다이오드**는 해외 수입 의존도가 높아질 것으로 예측됨

개선사항 및 기대효과

01 개선사항

- 코로나 19 장기화됨에 따라 수출입 금액의 변동이 불확실해진 가운데 **정교한 학습 필요**
→ 정부 차원 혹은 기업 차원에서 전폭적인 지원
→ 대응방안 모색에 더욱 유의미한 정보를 제공 필요
- 한국을 비롯한 주요국을 대상으로 반도체 및 자동차 부품 수출경쟁력 분석을 진행하였기에 **특정 시장에서만 국한됨**
→ 미국과 유럽 등 기술 선진국들이 대부분의 생산을 아시아 국가에 위탁(outsourcing)하고 있는 만큼 아시아간 수출 경쟁력 비교 필요
- 반도체 및 자동차 부품 산업 전망을 예측하기 위해 정교한 예측모델 정확도 개선을 위해 **많은 양의 데이터 필요**
→ 추후 뉴스 텍스트 데이터 등과 같은 **비정형 데이터를 계량화하여 새로운 정보 원천으로 활용** 필요

02 기대효과

- ① 시계열 분석을 통한 품목별 수출입 금액 변동 파악 및 예측 → 산업별 시장규모 파악 및 국가 경쟁력 전략 제고에 필요한 기초 자료 제공
- ② RCA와 TSI분석을 통해 경쟁력이 강한 산업군을 확인하여 정부차원에서 국가 간 무역 중개 사업을 진행 하였을 경우에 긍정적인 결과가 기대됨

Part 4

부록

- 모델링 후기 및 느낀점
- 참고문헌

모델링 후기 및 느낀점

이름	내용
김현아	회귀분석으로 예측 성능이 좋은 모델을 구현해냈듯이 더 다양한 모델링을 구축할 수 있었겠지만 시간과 데이터의 부족으로 진행하지 못했다는 아쉬움이 있습니다. 하지만 조원분들과 함께 의견을 나누고 배운 것을 활용해가며 프로젝트를 진행하는 과정에서 큰 보람을 느꼈습니다. 조원분들 모두 열정적으로 프로젝트에 임해 주셔서 저 또한 재밌게 참여했습니다.
박혜인	프로젝트를 진행하면서 하나하나 모델을 찾아가면서 공부를 하는 과정이 정말 유익했습니다. 주제를 정하는 데에 시간이 조금 오래 걸려서 마지막에 모델링을 할 시간이 조금 부족했지만, 여태 배웠던 내용을 활용하면서 큰 보람도 느끼고 즐거웠습니다.
임승찬	프로젝트를 구현하면서 팀원 간 각 담당의 업무 조율 및 피드백을 통해서 협업의 중요성이 자명할 수 있었습니다. 공통된 목표를 달성하기 위해서 창의적인 시도와 개선을 하여 새로운 가치를 만들어 개인적인 역량을 성장할 수 있었던 계기였습니다.
조완제	1차 프로젝트때와 비교해서 데이터 분석만 하기 때문에 훨씬 시간적 여유가 있을 것이라 생각했다. 막상 프로젝트를 시작해보니, 주제 선정부터 데이터 수집 및 전처리, 분석 까지 일련의 과정에 여유는 없었다. 각자 부족한 부분을 채워주면서 협력하는 팀프로젝트였다.

참고문헌

- [1] 김은영.(2021),한국과 주요국 간의 반도체산업 수출경쟁력 및 수출경합도 비교 분석
- [2] 산업통상자원부.(2022), 2021년 연간 수출입 동향
- [3] 라공우, 송진구.(2017), 미국시장에 대한 한·중 자동차부품산업의 국제경쟁력에 관한 연구
- [4] 김은영, 서창배.(2021), 한국과 주요국 간의 반도체산업 수출경쟁력 및 수출경합도 비교 분석
- [5] 이철,이광재.(2008), 자동차 부품산업의 국제경쟁력 요인과 수출성과와의 관계
- [6] 김영대, 윤보람.(2021) 수출전선 맑음 ... 반도체 호황기 추월, 연합마이더스 : 유가
- [7] 한국전자산업진흥회.(2005), 기계 전자 반도체 자동차 하반기 수출전망 '쾌청' - 물가, 환율, 유가
- [8] 모수원, 김창범.(2003), 철강제품과 반도체의 수출형태, 한국무역학회 세미나 및 토론회 - 환율
- [9] 손용정.(2001), 중 무역전쟁 전후 반도체. 디스플레이산업의 수출경쟁력 분석: 항공수출물류를 중심으로- 항공운임
- [10] 모수원, 김창범.(2003), 반도체와 선박의 수출형태, 국제무역연구 - 유가, 환율
- [11] 장선미.(2006), 한국 반도체산업의 무역구조와 국제경쟁력 분석, 산업경제연구 - 고용률
- [12] 산업연구원.(2020), i-KIET 산업경제이슈 : 코로나 19가 제조업 글로벌 공급망에 미치는 영향과 대응
- [13] 황선자, 이문호, 황현일.(2020), 자동차 산업의 구조변화와 정책과제 : 자동차부품산업을 중심으로
- [14] 금융감독원.(2019), 우리나라 수출에 영향을 미치는 주요 요인 (거시건전성감독국)

Thank You

질문이 있으신 분들은 질문 부탁드립니다 😊