

엑셀을 통한 빅데이터 분석 기초



I. 데이터 분석의 방향성



Value, 빠른 의사결정

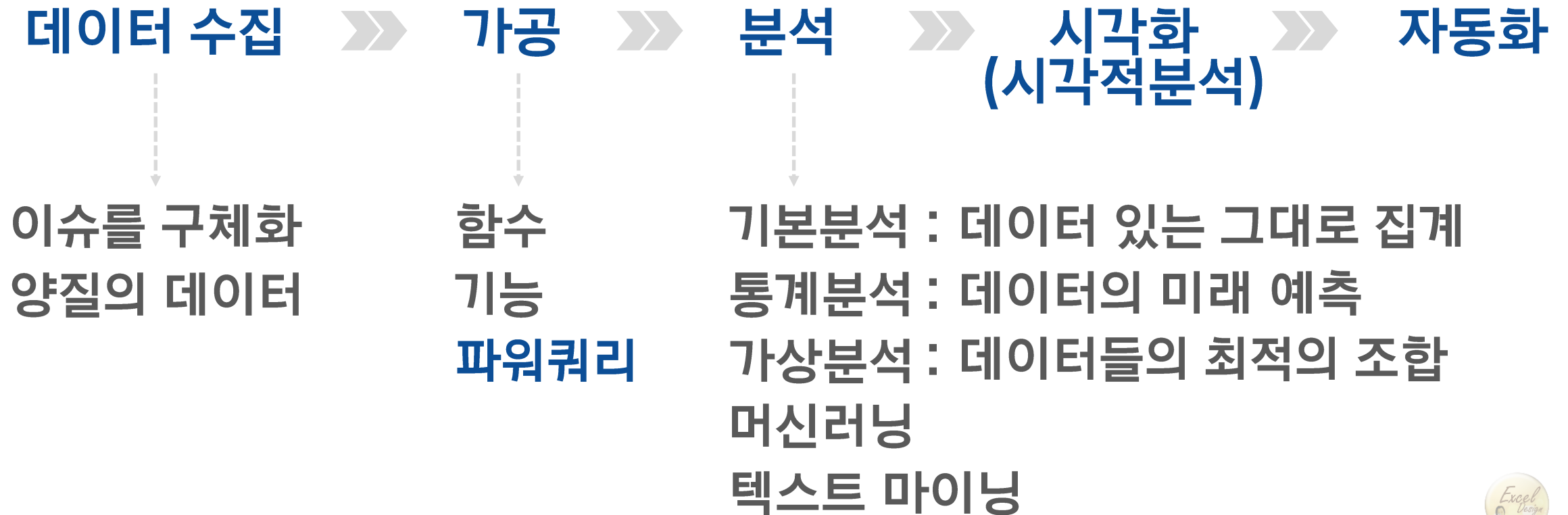


데이터 수집 / 가공 및 분석 / 시각화 / 자동화 역량



2. Data Handling Process

Data Science



II. Power Query 이해



3. 파워쿼리 개념 이해

1) 쿼리(Query)

- 특정한 조건을 만족하는 데이터만 추출해서 새로운 테이블 생성
- 고급필터의 개념과 비슷

2) 파워쿼리(Power Query)

- MS가 엑셀을 Business Intelligence Platform으로 만들겠다는 전략하에 나온 기능 중의 하나
- 파워BI = 파워쿼리 + 파워피벗 + 파워뷰 + 파워맵
- 2010~2013(추가 기능으로 설치), 2016버전 부터 기본 기능으로 안착
- 원본 데이터 보존, 쿼리 파일을 만들어 주는 방식, 언제든지 재가공 가능
- **새로고침으로 한방에 업데이트!!!**

원본 데이터 수정 → 새로고침 → 처음 설계한 편집 단계를 거쳐 → 자동 업데이트



4. 표 개체 이해

예제파일 : 표개체 이해.xlsx

- 1) 관련 데이터 그룹을 보다 쉽게 관리하고 분석하기 위해 셀범위를 Excel 표로 변환
- 2) 장점
 - 헤더의 드롭다운 목록을 사용하면 데이터를 쉽게 정렬하고 필터링 할 수 있다.
 - 레코드를 추가, 삭제 시 테이블 범위가 자동으로 확장 및 축소된다.
 - 수식 및 서식이 자동으로 채워진다.
 - 수식은 구조적참조를 사용하므로 처리속도가 빠르다.
 - 요약 행을 추가할 수 있다.
 - 피벗테이블로 기본분석이 쉬워진다.
 - 슬라이서를 추가할 수 있다.
- 3) 단점
 - 구조적참조를 사용하므로 셀 별로 수식을 다르게 지정할 수 없다
 - 표 아래 셀 잠금이 해제된 경우에는 보호된 시트에서 표가 자동으로 확장되지 않는다.
- 4) 표 개체 전환 프로세스
 - 일반 데이터범위 내, 셀 하나 클릭 > 리본 메뉴 > 삽입 > 표 (Ctrl + T) > 표 스타일 지정 > 표 이름 바꾸기
- 5) 표 개체를 일반 데이터범위로 바꾸기
 - 표 개체 내, 셀 하나 클릭 > 리본 메뉴 > 표 도구 > 테이블 디자인 > '범위로 변환' 클릭



5. 파워쿼리 가공 표준 프로세스

예제파일 : 파워쿼리 가공 스킬.xlsx

- 1) 일반 데이터 범위 → 표개체로 변환(Ctrl + T)
 - 파워쿼리로 가공하기 가장 좋은 형태
 - 원본 데이터 손상이 없음
 - ‘새로고침’ 클릭 한 번으로 가공작업 자동화
- 2) 변환된 표개체 내부 임의의 셀 클릭 > 리본메뉴 > 데이터 > 테이블/범위에서 > 파워쿼리 편집 창 > 데이터타입 변환, 열분할 > 홈 > 닫기 및 로드 > 새 시트, A1셀 부터 가공된 데이터 뿌려짐
- 3) 원본 데이터에 추가, 수정 발생 시 > 쿼리 테이블 내, 임의의 셀 클릭 > 우클 > ‘새로고침’ : 해당 쿼리 테이블만 update
- 4) 원본 데이터에 추가, 수정 발생 시 > 쿼리 테이블 내, 임의의 셀 클릭 > 리본메뉴 > 데이터 > ‘모두 새로고침’ : 해당 통합문서 내, 모든 쿼리 테이블 update



6. 데이터를 의심하고 검증하라

- 1) 정형 데이터 / 비정형 데이터
- 2) 데이터의 구조화(테이블화) : 구조화가 잘 된 테이블은 언제든지 데이터 분석을 빠르고 쉽게 진행 가능

Column1	Column2	Column3	Column4	Column5

- 3) 헤더(필드명)는 셀병합 없이 첫번째 행에만 입력
- 4) 아래 방향(↓)으로 데이터를 차곡차곡 쌓기
- 5) 중복이 발생해도 비워두지 말고 입력
- 6) 표 개체 여부
- 7) 필드별 데이터 타입 확인
 - 숫자 / 문자 / 날짜 / 수식

7. 파워쿼리 편집창 구조 / 데이터 타입 변환

The screenshot displays the Excel Power Query Editor interface. At the top, the ribbon menu is visible with tabs for '파일' (File), '홈' (Home), '변환' (Transform), '열 추가' (Add Column), and '보기' (View). The '변환' tab is active, showing options for data type changes like '데이터 형식: 임의' (Data Format: Arbitrary) and '첫 행을 머릿글로 사용' (Use first row as header). Below the ribbon, a data table is shown with columns: '날짜' (Date), '시간(시간단위)' (Time), '연령대(10세 단위)' (Age Group), '성별' (Gender), '시' (City), '군구' (District), and '유동인구 수' (Population). The '날짜' column is highlighted with a blue dashed arrow pointing to the '데이터 타입 바꾸기' (Change Data Type) icon in the ribbon. On the right, the '쿼리 설정' (Query Settings) pane is open, showing the '속성' (Properties) tab with the query name '표1' (Table1) and the '적용된 단계' (Applied Steps) list.

〈리본 메뉴〉

〈쿼리 이름〉

〈숫자〉

- 1.2 10진수
- \$ 통화
- 123 정수
- % 백분율
- 날짜/시간
- 날짜
- 시간
- 날짜/시간/표준 시간대
- 기간
- AB C 텍스트
- True/False
- 이진
- 로컬 사용...

〈쿼리표시 창〉

〈데이터 타입 바꾸기 : 아래 방법 중 택1〉

- 아이콘 클릭 > 데이터 타입 선택
- 헤더 선택 > 우클릭 > 형식 변경 > 데이터 타입 선택

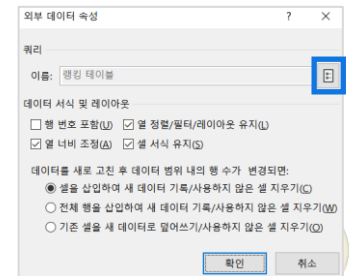
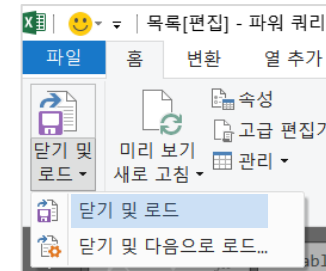
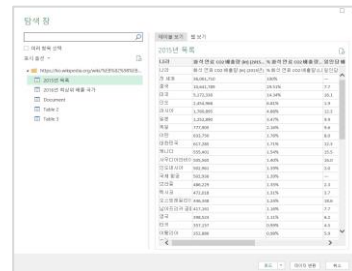
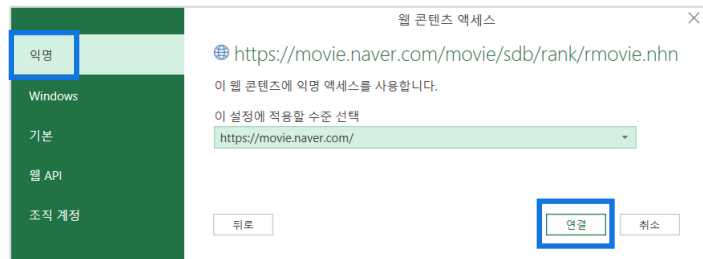
〈히스토리 창〉

III. 데이터 수집



8. 웹크롤링 1

- 1) 님스코리아 > 시청률 순위 > URL 복사
- 2) 새 통합문서 > 리본메뉴 > 데이터 > 새 쿼리 > 기타 원본에서 > 웹 > URL 붙여넣기
- 3) 좌측 탐색창에서 항목 선택 > 우측 미리보기 창에서 데이터 보임 > 원하는 항목 선택 > '데이터 변환' > 필요한 다양한 가공 진행
 - 첫 행을 머리글로 사용
 - 상위 행 제거 / 오류 제거 / 정렬
 - 데이터 타입 변환
- 4) 쿼리 편집창에서 데이터 가공 > 홈 > '닫기 및 로드'
- 5) 원본 데이터 수정되었을 시 > 리본메뉴 > 데이터 > '모두 새로고침'
- 6) 자동 업데이트 : 쿼리 테이블 클릭 > 리본 메뉴 > 데이터 > 속성 > 특정 시간 간격으로 자동 업데이트 가능

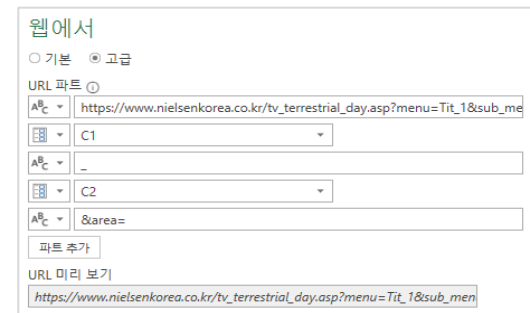


9. 웹크롤링 2

예제파일 : 닐슨시청률.xlsx

예제파일 : 닐슨시청률_URL.txt

- 1) '조건' 시트의 테이블들을 각각 '표 개체' 로 전환
- 2) 각각의 표 개체를 파워쿼리로 로드 > 모든 필드를 '텍스트' 타입으로 전환 > '번호' 필드 선택 > 리본 메뉴 > 변환 > 목록으로 변환 > 워크시트로 로드 시 > '연결 전용' 으로 로드
- 3) 엑셀 워크시트 리본 메뉴 > 데이터 > 데이터 가져오기 > 기타 원본에서 > 빈 쿼리
- 4) 파워쿼리 리본 메뉴 > 홈 > '매개변수 관리' > 새 매개변수 > 매개변수 설정 > 4개의 쿼리 모두 매개변수 설정
- 5) 위 4번 매개변수 쿼리들 모두 '연결 전용' 으로 로드
- 6) 엑셀 워크시트 리본 메뉴 > 데이터 > 데이터 가져오기 > 기타 원본에서 > 웹 > 고급 > 매개변수 파트 추가

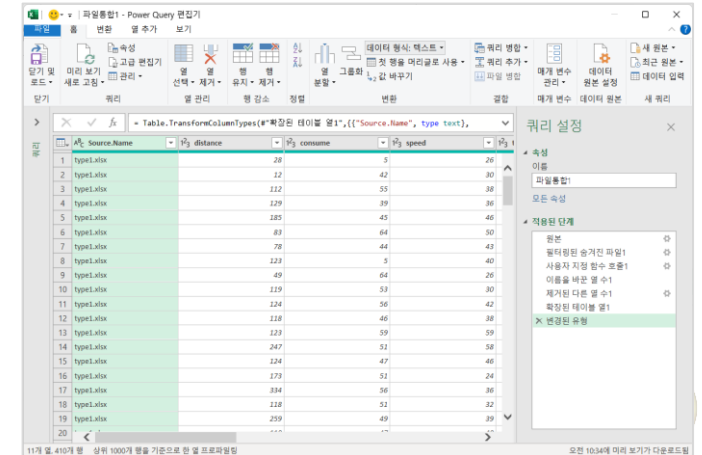
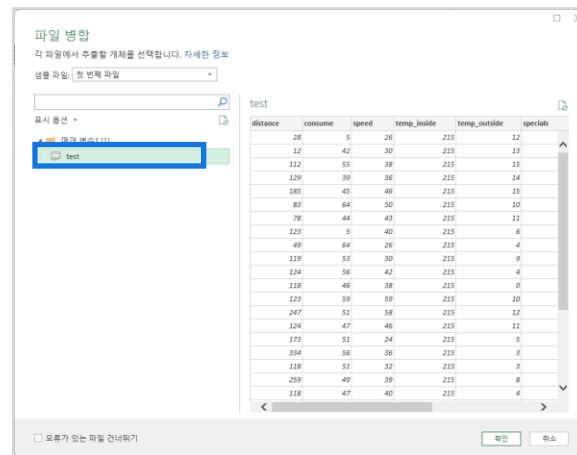
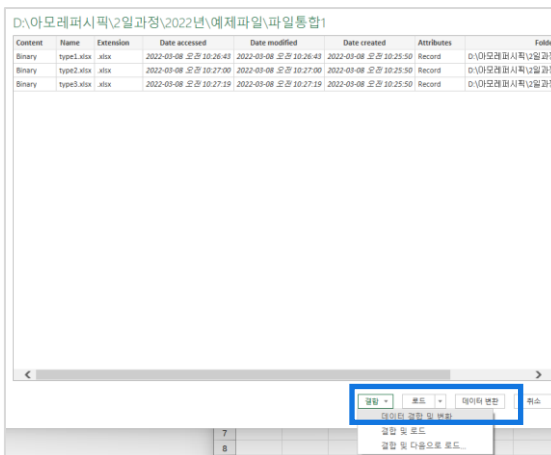


10. 특정 폴더 내, 엑셀 파일 통합 1

예제폴더 : 파일통합1

※ 폴더 내, 엑셀 파일만 존재 + 취합할 시트 이름 모두 같음 + 시트별 필드 이름, 순서, 개수 모두 같음
폴더 내, 엑셀 파일, test 시트 데이터 취합하기

- 1) 새 통합문서 > 리본 메뉴 > 데이터 > 데이터 가져오기 > 파일에서 > 폴더에서 > 파일 취합할 폴더 선택
- 2) 우측 하단 > '결합' 내림버튼 클릭 > '데이터 결합 및 변환' 클릭
- 3) 파일 병합 창에서 > 좌측 중간 > 시트명 선택 > 확인
- 4) 취합 완료 된 데이터 테이블 완성 > 좌측 첫번째 필드에 파일명이 입력된 필드 생성 됨 > 불필요시 열 제거
- 5) 파워쿼리 리본메뉴 > 홈 > 닫기 및 로드 > 워크시트로 쿼리 테이블 로드 완료
- 6) 업데이트 검증하기 : 폴더 내, 엑셀 파일을 추가 > 쿼리 테이블에서 '새로 고침' > 업데이트 완료



11. 특정 폴더 내, 엑셀 파일 통합2

예제폴더 : 파일통합2

※ 폴더 내, 다양한 파일 존재 + 취합할 시트 이름 모두 다름 + 시트별 필드 이름, 순서, 개수 모두 같음
폴더 내, 엑셀 파일, 모든 시트 데이터 취합하기

- 1) 새 통합문서 > 리본 메뉴 > 데이터 > 데이터 가져오기 > 파일에서 > 폴더에서 > 파일 취합할 폴더 선택
- 2) 우측 하단 > '데이터 변환' 클릭
- 3) 'Folder Path' 에서 하위폴더 Uncheck
- 4) 'Content' 필드를 제외한 나머지 필드들 열 제거 > '병합' 아이콘 클릭

	Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
1	Binary	type1.xlsx	.xlsx	2022-03-08 오전 11:19:03	2022-03-08 오전 11:19:03	2022-03-08 오전 10:36:26	Record	D:\아모레퍼시픽\2월과정\2022년\여
2	Binary	type2.xlsx	.xlsx	2022-03-08 오전 10:36:26	2020-10-18 오후 1:22:31	2022-03-08 오전 10:36:26	Record	D:\아모레퍼시픽\2월과정\2022년\여
3	Binary	type3.xlsx	.xlsx	2022-03-08 오전 10:36:26	2020-10-18 오후 1:23:28	2022-03-08 오전 10:36:26	Record	D:\아모레퍼시픽\2월과정\2022년\여

- 5) 파일 병합 창에서 > 좌측 중간 > '매개 변수' 선택 > 확인
- 6) 'Data' 필드를 제외한 나머지 필드들 열 제거 > '확장' 아이콘 클릭 > '원래 열 이름을 접두사로 사용' Uncheck
- 7) 1행 데이터를 헤더로 옮기기 : 파워쿼리 리본메뉴 > 홈 > '첫 행을 머리글로 사용'
- 8) 나머지 헤더들 필터링 처리하기 : 'distance' 필드 내림버튼 클릭 > 'distance' Uncheck
- 9) 파워쿼리 리본메뉴 > 홈 > 닫기 및 로드 > 워크시트로 쿼리 테이블 로드 완료
- 10) 업데이트 검증하기 : 폴더 내, 엑셀 파일을 추가 > 쿼리 테이블에서 '새로 고침' > 업데이트 완료



12. 시트통합 1

예제파일 : 시트통합1.xlsx

※ 특정 파일 내, 여러 시트 존재 + 취합할 시트 이름 모두 다름 + 시트별 필드 이름, 순서, 개수 모두 같음
파일 내, 모든 시트 데이터 취합하기

- 1) 새 통합문서 > 리본 메뉴 > 데이터 > 데이터 가져오기 > 파일에서 > 통합 문서에서 > 시트 취합할 파일 선택
- 2) 탐색 창에서 > 좌측 중간 > 파일명(시트통합1.xlsx) 선택 > 우측 하단 > 데이터 변환 클릭
- 3) 'Data' 필드를 제외한 나머지 필드들 열 제거 > '확장' 아이콘 클릭 > '원래 열 이름을 접두사로 사용' 해제

	AB C Name	Data	AB C Item	AB C Kind	Hidden
1	3월	Table	3월	Sheet	FALSE
2	4월	Table	4월	Sheet	FALSE
3	5월	Table	5월	Sheet	FALSE

- 4) 1행 데이터를 헤더로 옮기기 : 파워쿼리 리본메뉴 > 홈 > '첫 행을 머리글로 사용'
- 5) 나머지 헤더들 필터링 처리하기 : 'Date' 필드 내림버튼 클릭 > 'Date', '(Null)' Uncheck
- 6) 파워쿼리 리본메뉴 > 홈 > 닫기 및 로드 > 워크시트로 쿼리 테이블 로드 완료
- 7) 업데이트 검증하기 : 특정 파일 내, 시트 추가 > 쿼리 테이블에서 '새로 고침' > 업데이트 완료

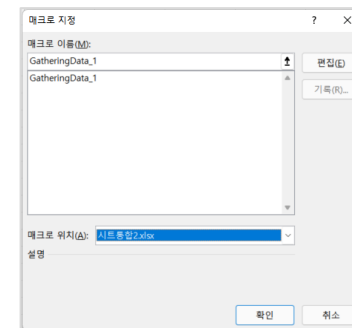
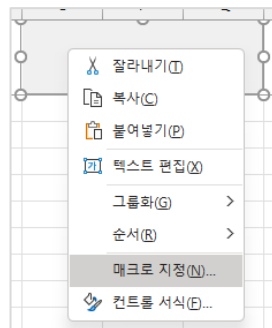


13. 시트통합2

예제파일 : 시트통합2.xlsx
패턴코드_시트통합.txt

※ 특정 파일 내, 여러 시트 존재 + 취합할 시트 이름 모두 다름 + 시트별 필드 이름, 순서, 개수 모두 다름
파일 내, 모든 시트 데이터 취합하기

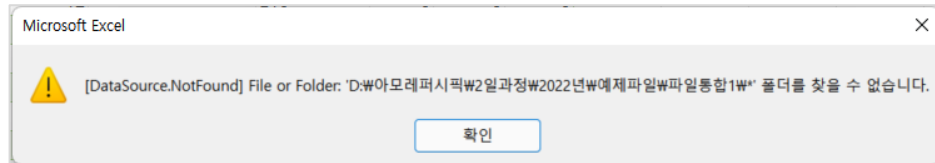
- 1) 시트통합2.xlsx 파일 열기 > Alt + F11 을 눌러 VBE 창 열기 > 좌측 프로젝트 탐색기 창에서 > 파일명(시트통합2.xlsx) 더블클릭 > 우클 > 삽입 > 모듈
- 2) 우측 코드 창에 > 패턴코드(패턴코드_시트통합.txt) 복사, 붙여넣기
- 3) 매크로버튼 > 우클 > '매크로 지정' 선택 > 프로시저 선택(GatheringData_1), 매크로 위치 선택(시트통합2.xlsx) > 확인
- 4) 리본 메뉴 > 파일 > 다른 이름으로 저장 > 파일 확장자를 '매크로 사용 통합문서' 로 변경해서 저장
- 5) 실무 적용 시
 - 시트 별, 필드 이름, 순서, 개수 원하는대로 편집, 추가
 - 코드 9번째 줄(시트 이름 출력 코드) 필요 없을 시, 주석처리(해당 코드 맨 앞에 작은따옴표(') 입력)



14. 원본 데이터 경로 이동 시, 업데이트

예제폴더 : 파일통합, 경로 이동
예제파일 : 파일통합 완성 파일 활용

1) 원본 폴더의 경로를 이동시키면 기존 쿼리에서 아래와 같은 에러 발생



2) 우측 '쿼리 표시' 창에서 맨 하단의 쿼리 더블클릭

3) 파워쿼리 리본 메뉴 > 우측 'History' 창에서 '원본' 스텝 선택

4) 파워쿼리 리본 메뉴 > 홈 > 데이터 원본 설정 > 원본 설정 창에서 > 경로 선택 > 좌측 하단 > '원본 변경' 클릭 > '찾아보기' 클릭 해서 이동 된 폴더 선택 > 확인

5) 파워쿼리 리본 메뉴 > 홈 > '미리보기 새로 고침' 내림버튼 클릭 > '모두 새로 고침' 클릭 > 업데이트 완료



IV. 데이터 가공

데이터 전처리 실무



15. 크로스테이블 구조화

예제파일 : 크로스테이블.xlsx

- 1) 데이터범위 내, 임의의 셀 클릭 > 리본 메뉴 > 데이터 > 테이블에서
- 2) 파워쿼리 편집 창 > Jan ~ Dec 필드 선택 > 우클 > 선택한 열만 피벗해제
- 3) Year, 특성 필드 선택 > 우클 > 열 병합 : 구분기호(사용자 지정 / '-')
- 4) '병합됨' 필드 > 우클 > 형식변경 > 날짜
- 5) 홈 > 닫기 및 로드 > 닫기 및 로드

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2	2012	11234	13910	19233	22311	22709	21543	20533	19399	18433	17135	15820	11823
3	2013	11022	12912	17887	21098	23102	22380	21229	20003	18538	15738	15511	10129
4	2014	10587	12491	18588	21718	22747	21083	20409	20154	17792	17123	14521	10897
5	2015	10064	12526	19389	21686	21670	17369	20104	19292	17994	6017	15102	10202
6	2016	10200	12587	18984	22314	22116	22741	21127	20093	18210	15846	15567	10046
7	2017	9527	11182	17448	21808	22948	20216	20004	19696	19830	16072	15611	11681
8	2018	8456	10232	16604	19906	21389	22094	21803	20724	17304	17131	15528	11218
9	2019	9965	11646	17793	20749	22229	21557	21193	21041	18271	18052	16443	12405
10	2020	11369	12108	8013	4867	13522	21093	21687	22195	20737	19813	15998	48716
11	2021	9115	12432	19667	20081	21738	22274	20943	20711	0	0	0	0

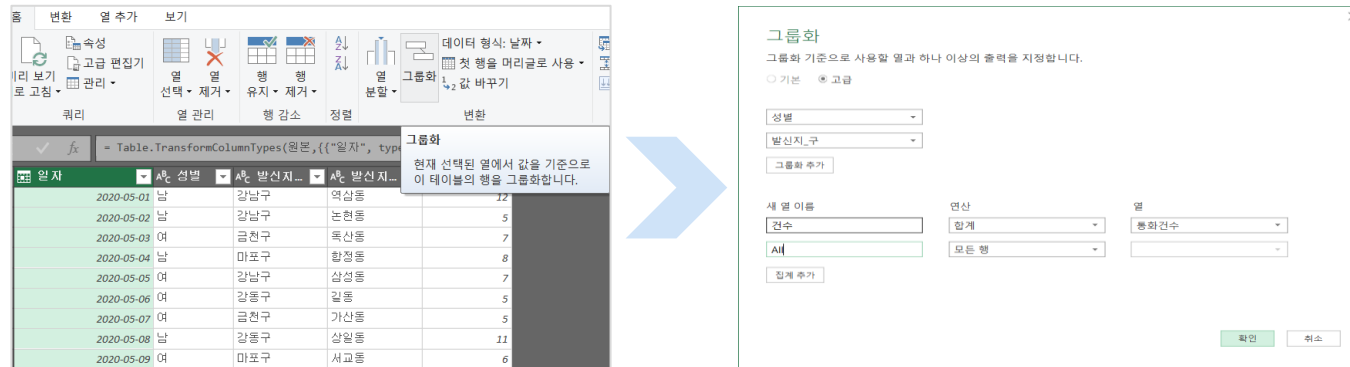
Date	Sales
2012-01-01	11234
2012-02-01	13910
2012-03-01	19233
2012-04-01	22311
2012-05-01	22709
2012-06-01	21543
2012-07-01	20533
2012-08-01	19399
2012-09-01	18433
2012-10-01	17135
2012-11-01	15820
2012-12-01	11823
2013-01-01	11022
2013-02-01	12912
2013-03-01	17887
2013-04-01	21098
2013-05-01	23102
2013-06-01	22380
2013-07-01	21229
2013-08-01	20003
2013-09-01	18538
2013-10-01	15738



16. 피벗테이블의 진화

예제파일 : 피벗테이블의 진화.xlsx

1) 홈 > 그룹화



그룹화

그룹화 기준으로 사용할 열과 하나 이상의 줄력을 지정합니다.

☐ 기본 ☒ 그룹

성별

발신지_구

그룹화 추가

새 열 이름

연산

열

연수

합계

통과연수

All

모든 행

그룹화 추가

확인 취소

2) 열 추가 > 사용자지정 열 > 수식 입력 > 'All' 필드 제거



사용자 지정 열

다른 열에서 계산된 열을 추가합니다.

새 열 이름

사용자 지정

사용자 지정 열 수식

=Text.Combine(List.Distinct([All][발신지_동]), ", ")

사용 가능한 열

성별

발신지_구

연수

All

<< 삽입

파워 쿼리 수식에 대해 알아보기

✓ 쿼문 오류가 검색되지 않았습니다.

확인 취소



17. 데이터 구조화 실무

예제폴더 : 데이터 구조화

1) 어느 기업의 매출, 판매수량, 지점에 관한 데이터 테이블들을

- ① 연결하여
- ② 분석하기 바람직하도록
- ③ 데이터를 구조화해보자

쿼리 및 연결

쿼리

연결

4개 쿼리

매출

연결 전용입니다.

판매수량

연결 전용입니다.

지점정보

연결 전용입니다.

AllData

239,952개의 행이 로드되었습니다.

적용된 단계

원본

탐색

변경된 유형

승격된 헤더

변경된 유형1

이름을 바꾼 열 수

제거된 상위 행 수

선택한 열만 피벗 해제됨

변경된 유형2

이름을 바꾼 열 수1

변경된 유형3

적용된 단계

원본

확장된 판매수량

병합된 쿼리

확장된 지점정보

다시 정렬한 열 수

	A	B	C	D	E	F	G	H	I
1	지점코드	지점명	시도	시군구	제품카테고리	제품명	기간	매출	판매수량.1
2	B1000	경기101	경기	부천시	A009	A009117	2018-01-01	4989269	9465
3	B1000	경기101	경기	부천시	A009	A009117	2018-02-01	399730	2737
4	B1000	경기101	경기	부천시	A009	A009117	2018-03-01	669579	4953
5	B1000	경기101	경기	부천시	A009	A009117	2018-04-01	-3835461	3877
6	B1000	경기101	경기	부천시	A009	A009117	2018-05-01	-4529788	3350
7	B1000	경기101	경기	부천시	A009	A009117	2018-06-01	4456445	6554
8	B1000	경기101	경기	부천시	A009	A009117	2018-07-01	2812640	9097
9	B1000	경기101	경기	부천시	A009	A009117	2018-08-01	2179480	3984
10	B1000	경기101	경기	부천시	A009	A009117	2018-09-01	1919284	7208
11	B1000	경기101	경기	부천시	A009	A009117	2018-10-01	-4211197	2736
12	B1000	경기101	경기	부천시	A009	A009117	2018-11-01	-5165215	7136
13	B1000	경기101	경기	부천시	A009	A009117	2018-12-01	-3158866	937
14	B1000	경기101	경기	부천시	A009	A009117	2019-01-01	2439069	3104
15	B1000	경기101	경기	부천시	A009	A009117	2019-02-01	-3086134	1472
16	B1000	경기101	경기	부천시	A009	A009117	2019-03-01	4160943	6994
17	B1000	경기101	경기	부천시	A009	A009117	2019-04-01	2306274	4924
18	B1000	경기101	경기	부천시	A009	A009117	2019-05-01	-2334602	8686
19	B1000	경기101	경기	부천시	A009	A009117	2019-06-01	-3445307	1756
20	B1000	경기101	경기	부천시	A009	A009117	2019-07-01	1262890	7236
21	B1000	경기101	경기	부천시	A009	A009117	2019-08-01	3041126	180
22	B1000	경기101	경기	부천시	A009	A009117	2019-09-01	3150993	4858
23	B1000	경기101	경기	부천시	A009	A009117	2019-10-01	1579849	3004
24	B1000	경기101	경기	부천시	A009	A009117	2019-11-01	2482956	6756

IV. 데이터 가공

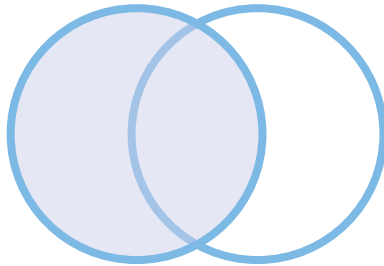
JOIN & UNION



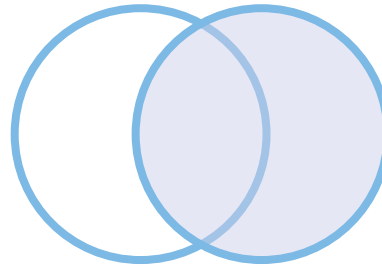
18. Data JOIN

예제파일 : Join.xlsx

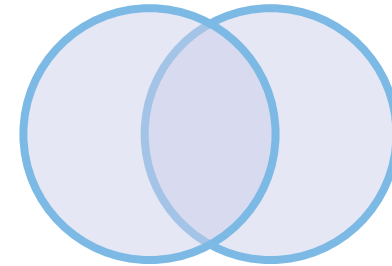
왼쪽 외부



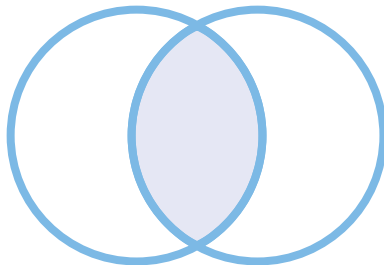
오른쪽 외부



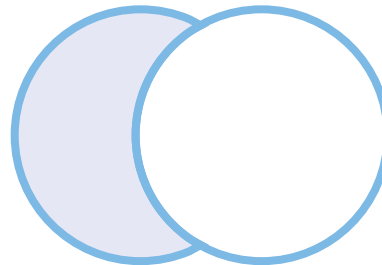
완전 외부



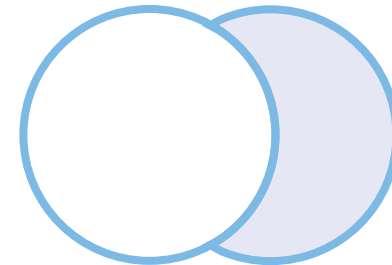
내부



왼쪽 안티



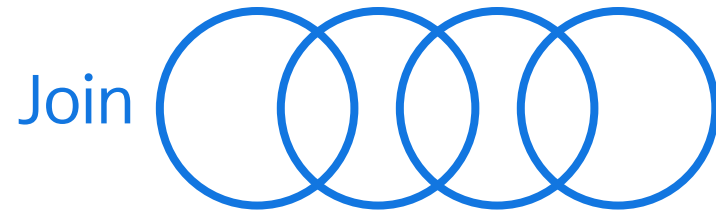
오른쪽 안티



19. Data UNION

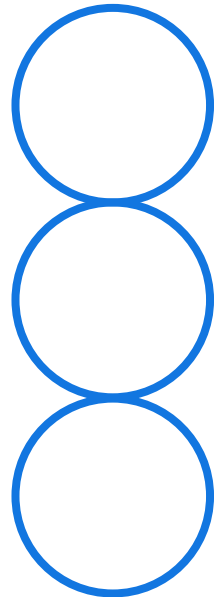
예제파일 : Union.xlsx

1) 병합(JOIN) : 데이터 테이블들을 옆으로 합치기



2) 추가(UNION) : 데이터 테이블들을 아래로 합치기

Union



병합

병합된 테이블을 만들 테이블 및 일치하는 열을 선택합니다.

Orders

site or Province	City	Postal Code	Order Date	Ship Date	Profit	Quantity ordered new	Sales
inois	Addison	60101	2012-05-28	2012-05-30	1.32	2	5.9
ashington	Anacortes	98221	2010-07-07	2010-07-08	4.56	4	13.01
ashington	Anacortes	98221	2011-07-27	2011-07-28	-47.64	7	49.92
ashington	Anacortes	98221	2011-07-27	2011-07-28	-30.51	7	41.64

Dates

Order Date	년	월의 시작
2010-01-01	2010	2010-01-01
2010-01-02	2010	2010-01-01
2010-01-08	2010	2010-01-01
2010-01-04	2010	2010-01-01
2010-01-05	2010	2010-01-01

조건 종류

[원래 외부(첫 번째의 모두, 두 번째의 일치하는 행)]

☐ 유사 일치를 사용하여 병합 비교

> 유사 항목 일치 옵션

확인 취소

Primary Key :
필드의 이름은 달라도
데이터 타입은 같아야 한다.

추가

세 개 이상의 테이블에 있는 행을 단일 테이블에 연결합니다.

☐ 2개의 테이블 ☒ 3개 이상의 테이블

사용 가능한 테이블

HK_US
HK_CA
HO_US
HO_CA
TE_US
TE_CA

추가 >>

추가할 테이블

HK_US
HK_CA
HO_US
HO_CA
TE_US
TE_CA

확인 취소



V. 데이터 분석

기본 분석 & 소셜 분석



20. 데이터 기본 분석

- 1) '인천데이터포털' 접속 > '교통사고' 키워드 검색 > ' 시도별 월별 교통사고' 클릭
- 2) KOSIS > '조회설정' > 데이터셋 다운로드
- 3) 파워쿼리로 데이터 가공 및 구조화
- 4) 피벗테이블로 기본분석
- 5) '조건부서식' 기능으로 시각화
- 6) 슬라이서 삽입으로 '동적 분석'
 - 유저가 원하는 조건으로 다각도 분석 가능
 - 인사이트를 가독성 높게 시각화

부상자수 (명)

사고건수 (건)

사망자수 (명)

합계 : 값2

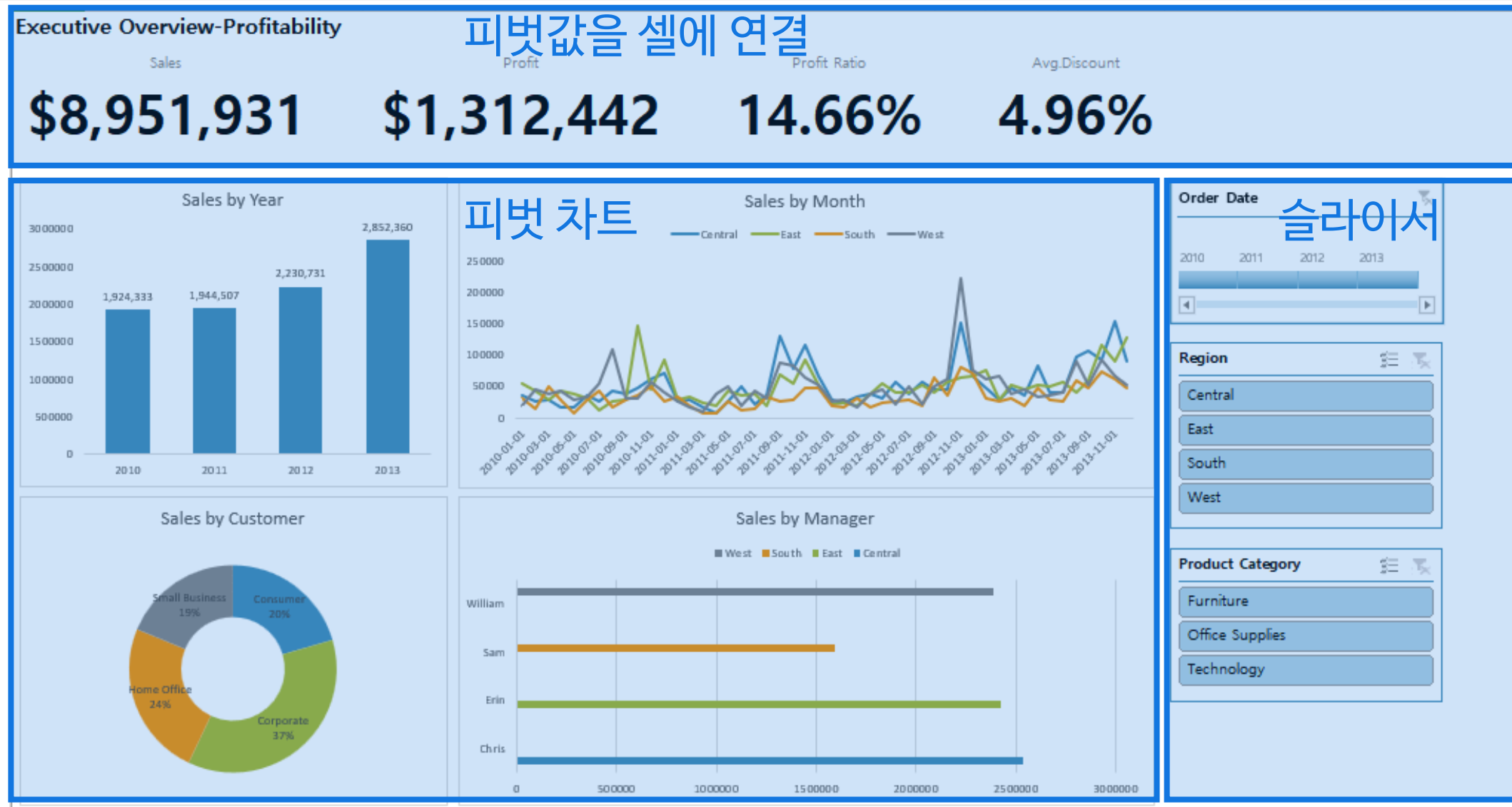
연

시도별(1)	2005년	2006년	2007년	2008년	2009년	2010년	2011년	2012년	2013년	2014년	2015년	2016년	2017년	2018년	2019년	2020년	2021년
강원	15634	15289	15918	15362	16560	15693	15630	15553	14505	14322	14987	14859	13532	11922	11752	10779	10221
경기	73705	70260	69395	65931	71011	70818	69141	71026	68430	73844	81763	78504	77551	81174	84508	77679	78225
강남	20637	20824	20783	21724	22878	21620	21067	21025	20274	19571	20141	18257	16838	16454	18563	16607	15414
경북	26277	26281	25027	24105	26098	26077	23773	24172	23319	23496	23880	22367	20707	20919	21970	18723	17690
광주	12134	12912	12794	13598	14761	14618	13467	13886	13089	13289	12596	11966	11876	11977	13076	12293	11489
대구	17957	19646	19896	20109	20858	21511	21479	20931	19713	20814	20433	18779	18553	19332	21598	18974	17230
대전	8857	8822	8664	8755	9697	9093	9172	8612	8234	8893	10413	11315	11798	11432	12649	10636	10175
부산	17829	19310	15078	17729	20263	19885	19816	21202	17542	18030	17884	16738	16157	16468	17832	16347	15457
서울	57434	56522	57517	59937	63584	59718	57625	58583	56761	57345	58656	55669	53810	53751	53904	47513	45223
울산	6576	7129	7239	7904	7899	7498	7400	7483	7305	7587	7943	6740	6127	5777	6301	5444	5238
인천	18467	17024	16305	16276	17074	15830	15418	14754	13594	13420	13124	12635	11257	10976	12258	11950	10730
전남	17575	18148	17857	18012	19103	18610	16991	17258	16808	16593	17020	15511	15672	15680	17905	15645	13522
전북	16476	15518	16643	15983	17244	16782	16514	15876	15017	15127	14089	12261	11948	10452	11651	9778	8906
제주	4916	5122	4769	4792	5676	5374	5108	5726	6415	6656	7142	6885	6602	6611	6818	6131	6725
충남	14893	14222	14845	14781	14803	15299	14612	13731	12736	12901	14515	13298	14158	13283	14087	13381	12197
충북	12866	13200	13176	13964	14366	14032	14178	14410	14222	14759	14927	15144	15136	15687	15539	13238	12135

VI. 시각적 분석



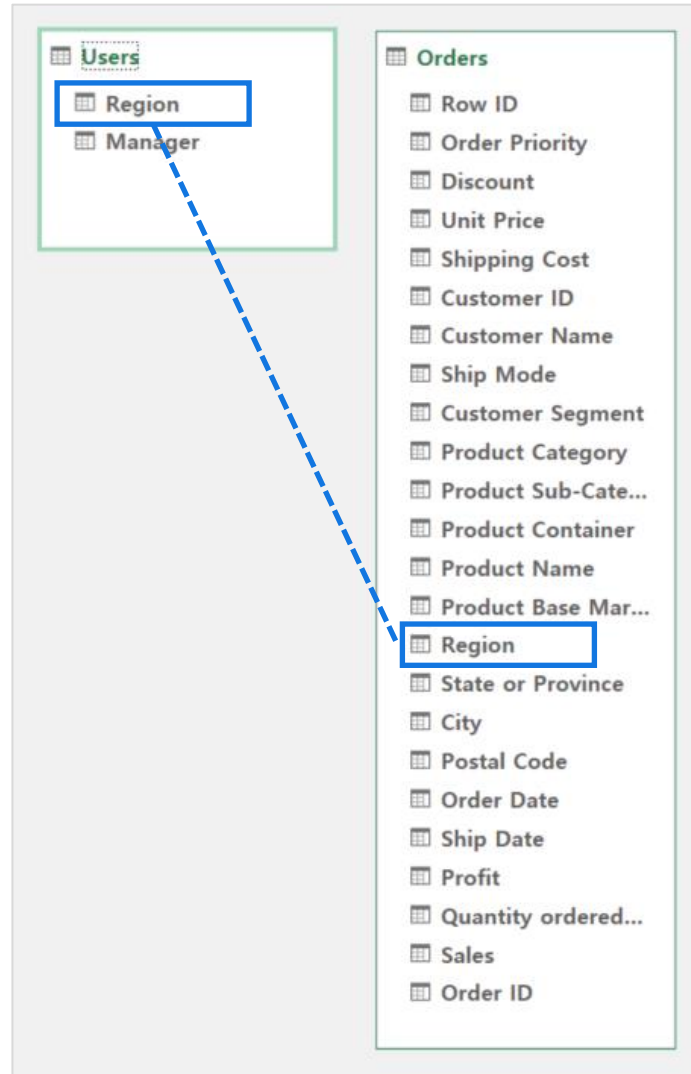
22. 대시보드 완성 미리보기



23. 데이터셋 구조

예제파일 : Dashboard.xlsx

1) 원본 데이터 구조



- 1:多 의 관계
- 고유목록:중복0
- 관련 필드끼리 데이터 타입이 같다



24. 데이터 연결 & 가공

- 1) Orders, Users 각 시트의 데이터 범위 내 임의의 셀 클릭 > Ctrl + T > 표 개체로 변환 > 표 이름 바꾸기 (표 이름 : Orders, Users)
- 2) 'Orders' 표 개체 내 임의의 셀 클릭 > 엑셀 리본메뉴 > 데이터 > 테이블에서 > 'Order Date', 'Ship Date' 필드 선택 > 데이터 타입 : 날짜 > 'Order Date' 필드 선택 > 파워쿼리 리본메뉴 > 열 추가 > 날짜 > 월 > 월의 시작
- 3) 파워쿼리 리본메뉴 > 홈 > 닫기 및 다음으로 로드 > '연결만 만들기' > 확인
- 4) 'Users' 표 개체 내 임의의 셀 클릭 > 엑셀 리본메뉴 > 데이터 > 테이블에서
- 5) 파워쿼리 리본메뉴 > 홈 > 닫기 및 다음으로 로드 > '연결만 만들기' > 확인

The image illustrates the process of connecting data in Excel. It shows a data table with columns like Region, State or Province, City, Postal Code, Order Date, and Ship Date. A dialog box for changing column format is visible. The '쿼리 설정' (Query Settings) pane shows the 'Orders' table. The '데이터 가져오기' (Get Data) dialog box shows the '연결만 만들기' (Only create connection) option selected. The final screenshot shows the Excel interface with the data connected.

26. 셀에 피벗 집계값 연결

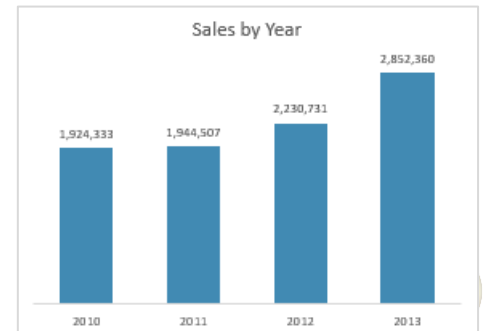
- 1) 'SalesDashboard' A4셀 클릭 > '=' 입력 > 마우스로 'Pivot' 시트의 A2셀 클릭 > Enter
- 2) 'SalesDashboard' D4셀 클릭 > '=' 입력 > 마우스로 'Pivot' 시트의 C2셀 클릭 > Enter
- 3) 'SalesDashboard' G4셀 클릭 > '=' 입력 > 마우스로 'Pivot' 시트의 E2셀 클릭 > Enter
- 4) 'SalesDashboard' J4셀 클릭 > '=' 입력 > 마우스로 'Pivot' 시트의 G2셀 클릭 > Enter

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Executive Overview-Profitability															
3		Sales		Profit		Profit Ratio		Avg.Discount								
4		\$8,951,931		\$1,312,442		14.66%		4.96%								

27. 피벗 차트 1

1. 연도별 판매 실적 차트(Sales by Year) 생성 프로세스

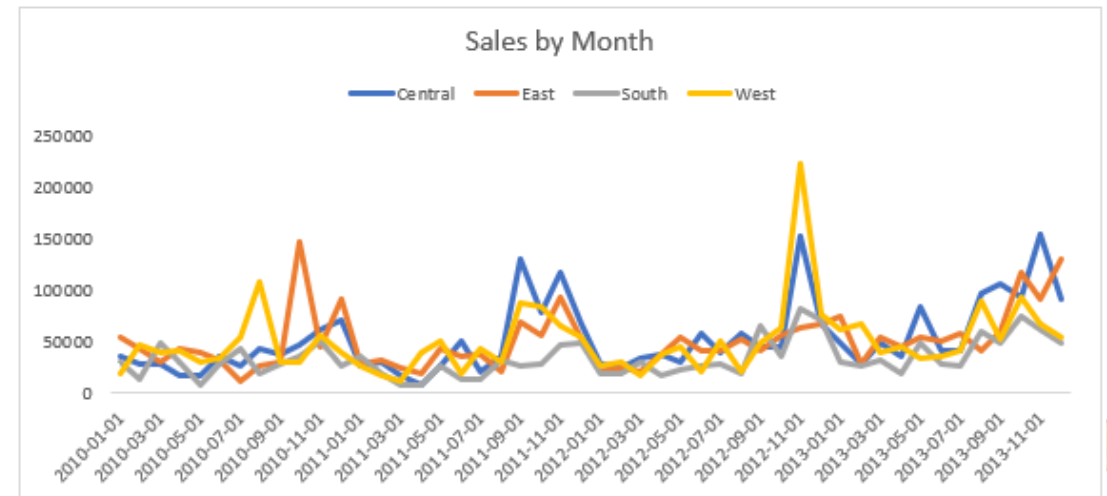
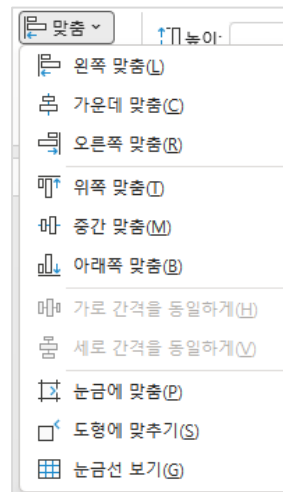
- 1) 엑셀 리본메뉴 > 삽입 > 피벗차트 > 대화상자에서 차트가 위치할 빈 셀 하나 클릭 > 확인
- 2) 'AllData' 테이블 객체 클릭 > 축 영역에 'Order Date', Σ 값 영역에 'Sales'
- 3) 차트가 선택된 상태에서 > 리본메뉴 > 피벗 차트 분석 > 필드 단추 > 모두 숨기기
- 4) 차트가 선택된 상태에서 > 오른쪽 > 차트 요소 아이콘 클릭(+) > 범례 해제, 눈금선 해제, 차트제목 편집
- 5) 차트 내, 데이터 계열 선택(파란색 bar) > 우클 > 데이터 레이블 추가
- 6) 추가된 데이터 레이블 숫자 클릭 > 우클 > 데이터 레이블 서식 > 표시 형식 > 범주 > 숫자 > '1000' 단위 구분 기호 사용' 체크
- 7) 차트 내, 데이터 계열 선택(파란색 bar) > 우클 > 데이터 계열 서식 > 계열 겹치기(0%), 간격 너비(100%)
- 8) 차트 일정 위치에 배치



28. 피벗 차트2

2. 월별 판매 실적 차트(Sales by Month) 생성 프로세스

- 1) 엑셀 리본메뉴 > 삽입 > 피벗차트 > 대화상자에서 차트가 위치할 빈 셀 하나 클릭 > 확인
- 2) 'AllData' 테이블 객체 클릭 > 범례 영역에 'Region', 축 영역에 '월의 시작', Σ 값 영역에 'Sales'
- 3) 차트가 선택된 상태에서 > 리본메뉴 > 피벗 차트 분석 > 필드 단추 > 모두 숨기기
- 4) 차트가 선택된 상태에서 > 오른쪽 > 차트 요소 아이콘 클릭(+) > 눈금선 해제, 범례 위쪽, 차트제목 추가
- 5) 차트가 선택된 상태에서 > 우클 > 차트 종류 변경 > 꺾은선형 > 확인
- 6) 차트의 사이즈 및 위치 조정 : 차트 선택 > 리본메뉴 > 도형 서식 > '맞춤' 아이콘으로 차트 위치 맞추기
- 7) 차트 일정 위치에 배치



29. 피벗 차트3

3. 고객별 판매 실적 차트(Sales by Customer) 생성 프로세스

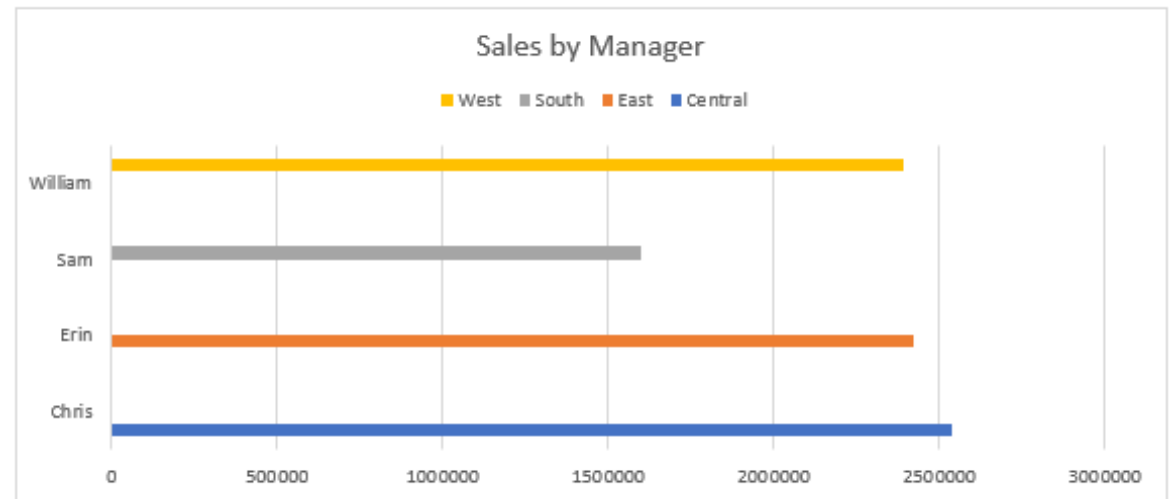
- 1) 엑셀 리본메뉴 > 삽입 > 피벗차트 > 대화상자에서 차트가 위치할 빈 셀 하나 클릭 > 확인
- 2) 'AllData' 테이블 객체 클릭 > 축 영역에 'Customer Segment', Σ 값 영역에 'Sales'
- 3) 차트가 선택된 상태에서 > 리본메뉴 > 피벗 차트 분석 > 필드 단추 > 모두 숨기기
- 4) 차트가 선택된 상태에서 > 오른쪽 > 차트 요소 아이콘 클릭(+) > 범례 해제, 눈금선 해제, 차트제목 편집
- 5) 차트가 선택된 상태에서 > 우클 > 차트 종류 변경 > 원형 > 도넛형 > 확인
- 6) 차트 내, 데이터 계열 선택 > 우클 > 데이터 레이블 추가 > 레이블 옵션 : 항목 이름+백분율
- 7) 차트의 사이즈 및 위치 조정 : 차트 선택 > 리본메뉴 > 도형 서식 > '맞춤' 아이콘으로 차트 위치 맞추기
- 8) 차트 일정 위치에 배치



30. 피벗 차트4

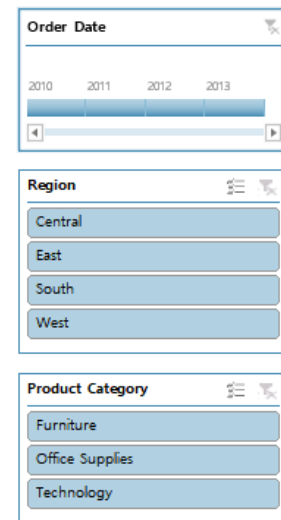
4. 매니저별 판매 실적 차트(Sales by Manager) 생성 프로세스

- 1) 엑셀 리본메뉴 > 삽입 > 피벗차트 > 대화상자에서 차트가 위치할 빈 셀 하나 클릭 > 확인
- 2) 'AllData' 테이블 객체 클릭 > 범례 영역에 'Region', 축 영역에 'Manager', Σ 값 영역에 'Sales'
- 3) 차트가 선택된 상태에서 > 리본메뉴 > 피벗 차트 분석 > 필드 단추 > 모두 숨기기
- 4) 차트가 선택된 상태에서 > 오른쪽 > 차트 요소 아이콘 클릭(+) > 범례 위쪽, 차트제목 추가
- 5) 차트가 선택된 상태에서 > 우클 > 차트 종류 변경 > 가로막대형 > 확인
- 6) 차트의 사이즈 및 위치 조정 : 차트 선택 > 리본메뉴 > 도형 서식 > '맞춤' 아이콘으로 차트 위치 맞추기
- 7) 차트 일정 위치에 배치



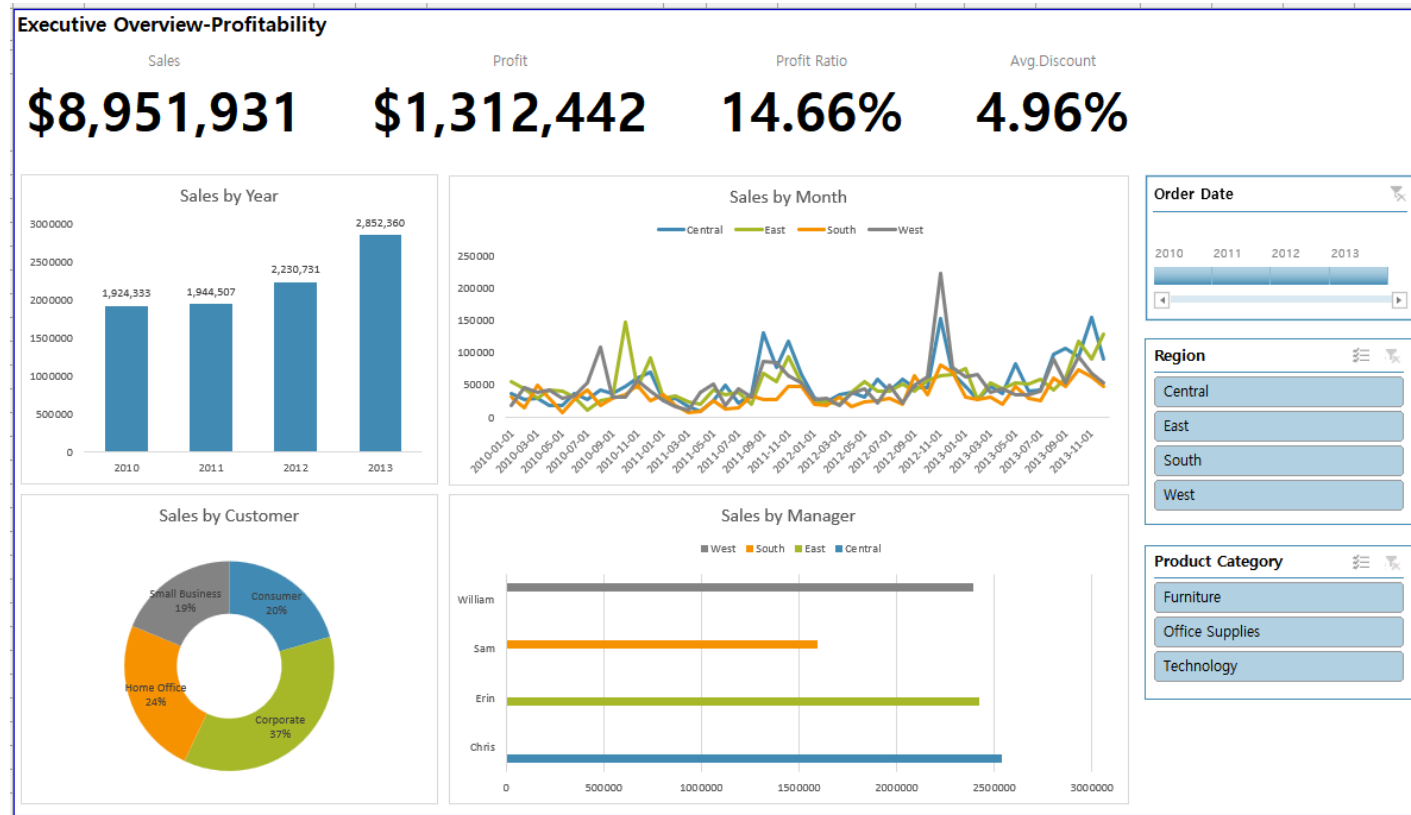
31. 슬라이서로 동적 분석

- 1) 'SalesDashboard' 시트, 빈 셀 클릭 > 리본 메뉴 > 삽입 > 필터 > 시간 표시 막대 > '기존연결' 대화상자 > 데이터 모델 탭 선택 > 이 통합 문서 데이터 모델 하나가 하이라이트 되어 있는 상태에서 > 열기 > 'Order Date' 체크 > 확인 > '모든 기간' → '년' > 리본 메뉴 > 타임라인(or 서식) > 표시 그룹 > '선택 레이블' 과 '시간 수준' 체크 해제
- 2) 'SalesDashboard' 시트, 빈 셀 클릭 > 리본 메뉴 > 삽입 > 필터 > 슬라이서 > 기존연결 대화상자 > 데이터 모델 탭 선택 > 이 통합 문서 데이터 모델 하나가 하이라이트 되어 있는 상태에서 > 열기 > 'Region', 'Product Category' 체크 > 확인
- 3) 'Order Date' 슬라이서 > 우클 > [보고서 연결](#) > 연도별 차트를 제외한 7개 개체에 모두 체크 > 확인
- 4) 'Region', 'Product Category' 슬라이서 > 우클 > [보고서 연결](#) > 8개 모든 개체에 모두 체크 > 확인



32. 깔끔한 인쇄 설정

- 1) 'SalesDashboard' 시트 > 리본 메뉴 > 보기 > 페이지 나누기 미리보기
- 2) 리본 메뉴 > 페이지 레이아웃 > 너비: 1페이지, 높이: 1페이지 > 용지방향: 가로 > 여백: 좁게 > 페이지 설정 옵션 버튼 > 여백 탭 > 가로, 세로 체크 > 인쇄 미리보기

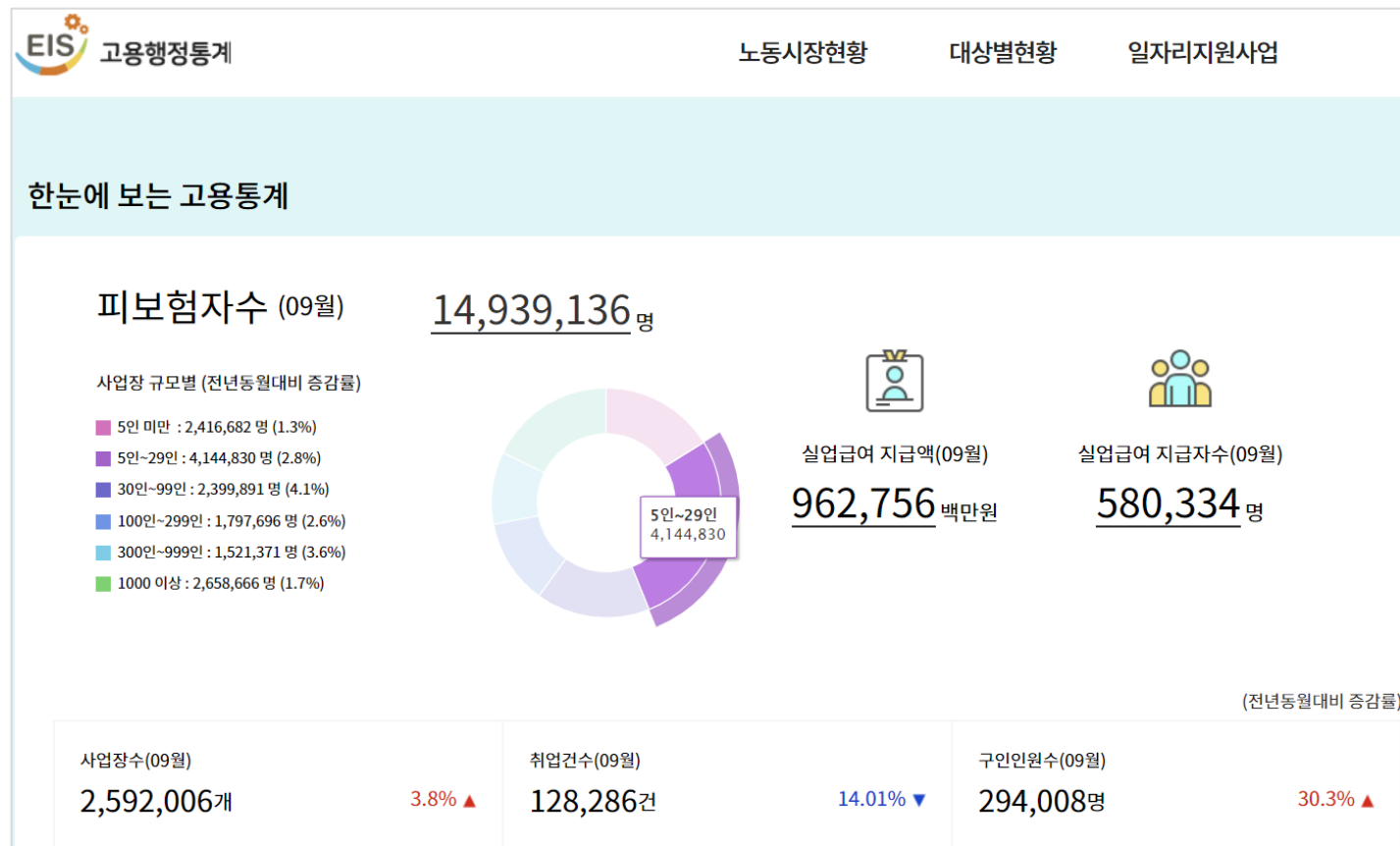


VII. Open API 활용 분석



33. 공공데이터 & Open API 활용 분석

- 1) 경기도 군포시, 22년 8~9월, 25~29세, 남녀별
- 2) 실업급여지급 합계액, 실업급여 최대값, 실업급여 지급건수를 분석하라



VIII. 통계 분석



엑셀

통계분석의 알고리즘은

통계적 모델에 기반을 두고 설계되었으므로
기본적 통계 지식이 있어야 한다.

1. 통계 : '현상'을 나타내고 있는 data로 부터 유용한 정보를 도출하여 적절한 의사결정을 하게 해주는 것

모집단/전수조사

대한민국
국민

표본/표본조사

1000명

유의수준(오차한계) 5%
신뢰수준 95%

1000명을 대상으로 여론조사 실시
47.2%가 특정당을 지지한다는 결론
대한민국 국민은 47.2%가 보수당을 지지한다...라고 표현
여기에는 오차가 발생할 확률이 있는데
이 오차가 일정한 한계를 넘지 않으면 좋겠다.
정확한 표현은
유의수준 5% 범위 내에서,
대한민국 국민은 47.2%가 보수당을 지지한다...라고 표현

2. 표본으로 전체를 설명하는 것을 ‘검정 통계학’

가설 검정 { 귀무 가설 : 기존의 사실과 차이가 없다, 같다
대립 가설 : 귀무 가설의 반대

유의수준(오차한계) 5%

P-Value < 0.05

귀무가설 기각, 대립가설 채택

-1 < 상관계수(ρ) < 1

- 1.0 : 완전한 음의 직선 상관관계
- 0.8 : 강한 음의 상관관계
- 0.3 : 약한 음의 상관관계
- 0 : 직선 상관관계가 아니다
- 0.4 : 약한 양의 상관관계
- 0.9 : 강한 양의 상관관계
- 1.0 : 완전한 양의 직선 상관관계

34. 기술통계법으로 데이터 속성을 한 눈에 파악

예제파일: 기술통계법.xlsx

1) 각 필드별 기술통계량 구하기 : 리본 메뉴 > 데이터 > 데이터분석 > '기술 통계법' 시행

	만족도 A	만족도 B
평균	79	79.1
표준 오차	3.3102711	1.52159888
중앙값	83	80
최빈값	88	84
표준 편차	14.8039824	6.80479707
분산	219.157895	46.3052632
첨도	-0.19116359	-0.18841956
왜도	-0.51478563	-0.66329701
범위	52	25
최소값	48	65
최대값	100	90
합	1580	1582
관측수	20	20

- 평균 비교 :
A와 B의 평균은 크게 차이가 없음
- 중앙값, 최빈값, 최소값, 최대값 비교
- 데이터의 흩어진 정도 비교 :
표준편차, 분산, 범위
 - ① A가 B보다 수치 높음
 - ② A에 비해 B가 중앙에 데이터가 모여 있음
 - ③ A는 B에 비해 중심에서 데이터가 멀어져 있음



35. 단순 회귀분석(Liner Regression)

예제파일 : 단순회귀분석.xlsx

- 1) **산포도 작성** : 상관 유무를 시각화
- 2) **상관분석** : 상관계수 파악 (**※상관관계와 인과관계는 다르다**)
- 3) **회귀분석** : 결정계수(회귀계수) 파악, 0~1 사이의 값, 클 수록 믿을만하다.
ex) y의 변화량 중 x의 변화로 설명할 수 있는 부분이 93%이고,
나머지 7%는 다른 요인에 의해 값이 변화한다고 해석할 수 있다.
- 4) **회귀식 파악(회귀 모형의 적합성 검정, $y=ax + b$)**
- 5) **유의한 F값 파악** : 회귀식을 통한 y의 값이 얼마나 정확한지 말해 준다.
0.05보다 작으면 해당 회귀식이 종속변수를 예측하는데 도움이 된다고 해석
- 6) **P-Value** :
0.05보다 작으면 해당 독립변수가 종속변수를 예측하는데 도움이된다고 해석
- 7) **잔차분석** : 잔차의 합계가 0인지 파악
- 8) **표준잔차** : 잔차를 표준편차로 나눈 값, 극단값을 찾는데 유용
표준잔차의 절대값이 3보다 크면 '이상치' 로 간주
이상치가 데이터 수집에서 오류가 있는 경우라면 이상치를 버리고 회귀 실행
수집 오류가 아니라면, 이상치 포함해서 회귀 실행, 이상치 제외하고 회귀 실행, 보고서에 두 결과를 모두 제시
- 9) **정규확률도** : 정규확률도가 직선에서 크게 이탈되면 정규분포 가정에서 어긋난다(잔차의 정규성 검정)
- 10) **잔차도** : 잔차의 등분산성, 잔차는 독립적이어야 한다, 잔차도에서 패턴이 없어야 한다.
- 11) **회귀식 확정**
- 12) **추세선 그리기**



36. 다중 회귀분석(Multiple Liner Regression)

예제파일 : 다중회귀분석.xlsx

1) 다중 독립변수 선택시 중요 규칙

- ① 종속변수에 대하여 영향력이 큰 독립변수를 찾는다
- ② 독립변수들 사이에 다중 공선성 (Variance Influence Factor:VIF)이 존재하면 인자 제거 후 분석
- ③ 가능한 한 최소의 변수로 최대의 효과를 본다
- ④ 결정계수는 클수록 좋다.
- ⑤ 결정계수와 조정된 결정계수와의 차이가 작을수록 좋다.

회귀 분석

입력

Y축 입력 범위(Y): \$C\$19:\$C\$32

X축 입력 범위(X): \$A\$19:\$B\$32

☒ 이통표(L) ☐ 상수에 0을 사용(Z)

☐ 신뢰 수준(D) 95 %

출력 옵션

☒ 출력 범위(O): \$H\$25

☐ 새로운 워크시트(P):

☐ 새로운 통합 문서(W)

잔차

☐ 잔차(R) ☒ 잔차도(D)

☒ 표준 잔차(I) ☐ 선적합도(I)

정규 확률

☒ 정규 확률도(N)

확인 취소 도움말(H)

	X1	X2	X3	X4	Y
X1	1				
X2	0.228579	1			
X3	-0.82413	-0.139242376	1		
X4	-0.24545	-0.972954999	0.029537	1	
Y	0.734494	0.813668069	-0.53582	-0.81933	1

