

# 회의록

## 1. 회의개요

일시	2021.07.30 14:00 ~ 17:00	장소	601강의실
작성자	최혜정	작성일	2021.07.30
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	주제 선정 및 활용 리소스 탐색		

## 2. 회의내용

<b>회의내용</b>	<p>1. 주제 선정</p> <ul style="list-style-type: none"> <li>- 프로젝트 주제 정하기 위한 탐색 (의료데이터와 라이프스타일을 이용한 질병 발병 위험도 예측 데이터셋/ 교통사고 분석 후 시각화 및 네비게이션 탑재/ 기상청 기온 자료와 전력소비데이터를 분석, 연간 기온에 따른 전력수요를 예측하여 에너지 부족상황에 대비 등 다양한 의견 도출)</li> </ul> <p>2. 주제 선별</p> <ul style="list-style-type: none"> <li>- 주제 1: 기온 자료와 전력 소비데이터, 신재생에너지 발전량 데이터를 바탕으로 (이상기후, 전기차 수요 증가 등으로 인한) 필요 전력량 충족 및 원활한 공급을 위한 솔루션 마련</li> <li>- 주제 2: 아동학대 사례분석 데이터, 한국 아동패널 특성 통계 정보를 바탕으로 데이터셋을 구축하고 인사이트를 통해 아동학대예방 솔루션 마련</li> </ul> <p>3. 데이터 조사 작업</p> <ul style="list-style-type: none"> <li>- 필요한 데이터 수집</li> <li>- 사용할 수 있는 데이터의 종류와 상태 확인</li> </ul>
<b>피드백</b>	<ul style="list-style-type: none"> <li>- 데이터를 설명할것</li> <li>- 왜 이 주제가 필요한지 객관적으로 설명하는 데이터가 필요</li> <li>- 쓰임새가 있고 기대치가 있는 주제 정할 것</li> </ul>
<b>결정사항</b>	<ul style="list-style-type: none"> <li>- 데이터 더 탐색하기</li> <li>- 2가지 주제 중 탐색하기 좋은 것이라고 판단되는 주제 선정해서 value와 인사이트 정리하기</li> <li>- 데이터 한 번씩 더 읽기</li> </ul>
<b>향후일정</b>	2차회의 : 2021.07.31, 강남역 이지스터디
<b>특이사항</b>	

## 3. 진행일지

계획	진행	완료	추후 계획
주제 탐색 및 데이터 탐색	주제 및 데이터탐색 -> 자료공유 -> 주제선별	주제선별완료	주제 선정 및 역할 분배, 데이터 전처리 계획, 인사이트 설정

# 회의록

## 1. 회의개요

일시	2021.07.31 18:00 ~ 20:00	장소	이지스터디
작성자	최혜정	작성일	2021.07.31
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	주제 선정 및 역할 분배		

## 2. 회의내용

<b>회의내용</b>	<p>1. 주제 선정</p> <ul style="list-style-type: none"> <li>- 전력 vs 아동, 주제별 탐색 정보 공유</li> <li>- value 비교(아동학대 예측 필요성, 전력 수요 예측 필요성)</li> </ul> <p>-&gt; 아동학대 관련 데이터 부족, 2011년 예비전력 부족으로 정전사태의 피해와 이상기후 및 전기차 수요증가로 인한 전력 사용량 증가의 이유로 전력 수요 예측 을 위한 데이터셋 구축으로 목표 설정.</p> <p>2. 자료 탐색</p> <ul style="list-style-type: none"> <li>- 데이터 수집 단위와 범위 결정 : 월별로 2016~2020 자료 수집</li> <li>- 예상 변수 : 기상정보, 경제성장률, 산업동향, 발전량, 전력소비량</li> <li>- 기대효과 : 정전피해예방, 수요관리 비용 절감, 재생에너지 공급 방향 설정 등</li> </ul> <p>3. 필요데이터 수집 역할 분담</p> <ul style="list-style-type: none"> <li>- 경제성장률(임희진), 기상(배송이), 산업동향(우동주), 전기차(최혜정)</li> </ul>
<b>결정사항</b>	- 각자 맡은 데이터 수집 및 정리 해오기
<b>향후일정</b>	3차회의 : 2021.08.02, 이대역 그레이프라운지
<b>특이사항</b>	

## 3. 진행일지

작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	1) 전기차 수요 증가 현황 조사 2) 전기차 등록 현황, 충전소 데이터 수집, 정리	데이터 탐색	* openurl 가져오기(urllib 이용) * 친환경 내수 데이터 수집
임희진	1) 경제 종합지수 2) 경제 분석지표 3) 부문별 전력 사용량 4) 에너지 소비량 데이터 수집 및 도메인 데이터 정리	데이터 탐색	* openurl 데이터 수집(requests 이용) * 수집 데이터 csv로 저장(pandas 이용)
배송이	1) 지역별 기상데이터 수집 2) 기상데이터에 들어가는 세부요소 확인 3) 기상과 전력소비에 관한 선행연구 조사	데이터 탐색	* openurl 가져오기(requests 이용)
우동주	1) 산업현황/산업단지/공단 조성 현황 조사 2) 산업체에서 사용하는 전력량 조사	데이터 탐색	* pandas read_excel

# 회의록

## 1. 회의개요

일시	2021.08.02 14:00 ~ 18:00	장소	그레이트라운지
작성자	최혜정	작성일	2021.08.02
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	수집한 데이터의 탐색 (목적에 맞게 데이터간의 연관성이 보이는지 확인)		

## 2. 회의내용

회의내용	<p>1. 데이터셋 방향 설정</p> <ul style="list-style-type: none"> <li>- 종속변수: 월별 전력 사용량 예측</li> <li>- 독립 변수: 가정용, 공공용, 산업용, 기타용 전력 사용량</li> </ul> <p>가정용 -&gt; 월별 기온에 따라 가정용 전력량 변동 공공용 -&gt; 전기차 증가에 따라 공공용 전력량 변동 산업용 -&gt; gdp 및 산업단지 조성에 증감에 따라 산업용 전력량 변동 기타용 -&gt; 고정 비율</p> <p>각 데이터별 상관관계 확인 후 카테고리화 진행 해야함.</p> <p>2. 자료의 확인 및 문제점 발견</p> <ul style="list-style-type: none"> <li>- 구축할 데이터셋에 맞는 자료인지 확인</li> <li>- 기대했던 인사이트에 맞는 데이터셋이 구축되는지 점검</li> <li>-&gt; 특징이 뚜렷하게 나타나지 않아 분석하기 어려운 문제가 있음.</li> </ul> <p>3 방향의 변경과 변수 확인.</p> <ul style="list-style-type: none"> <li>- 7, 8월달의 5년간 일별 기온과 전력 수요량을 이용하여 여름철 예비전력의 적정량을 예측하는 것이 어떨지 범위가 좁혀짐.</li> <li>- 종속변수 : 최대전력</li> <li>- 독립변수 : 기온, 습도 등 기온데이터, 체감기온, 불쾌지수 등</li> <li>- 파생변수 : 금년 전력 사용량*가정용 점유율(%) (가정에서 사용되는 전력 사용의 변화가 있는지)</li> </ul>
피드백	<ul style="list-style-type: none"> <li>- 주제를 완전히 바꾸지 말고 더 구체화 하고 데이터를 추가할 필요가 있음.</li> <li>- 전력 공급과 관련된 다른 경제지표 인사이트를 발견 할 수 있는지 찾아볼 것.</li> <li>- 전력 사용의 효율을 도출 가능 할 수 있을지, 에너지 효율 측면에 value를 낼 것.</li> <li>- 전력과 사람과 연결되는 것이 있을지 생각 해보기.</li> </ul>
결정사항	<ul style="list-style-type: none"> <li>- 날씨 open Api로 각자 맡은 범위의 데이터 수집해오기.</li> <li>- 카테고리 범주화 또는 어떤 인사이트가 도출되기를 기대하는지 대략적으로 구상해오기.</li> </ul>
향후일정	4차회의 : 2021.08.03, 이대역 그레이트라운지
특이사항	

## 3. 진행일지

작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	1) 2016년~2020년 7,8월 최대전력, 공급 예비력, 예비율 데이터 정리	데이터 수집	판다스를 이용한 데이터프레임 정리. 파일읽기 및 저장
임희진	1) 프로젝트 기획안 작성 및 계획 재정립 2) 날씨 데이터의 상관관계 확인 후 인사이트 설정	계획 수립	
배송이	1) 2016~2020년 날씨 데이터 저장 2) 전력량에 영향을 미치는 날씨데이터 변수 설정	데이터 수집 및 탐색	API 데이터 for문으로 불러오기
우동주	1) 국내 공단 내 입주 기업 수 조사	데이터 수집	데이터 프레임 조작

# 회의록

## 1. 회의개요

일시	2021.08.03 14:00 ~ 18:00	장소	그레이트라운지
작성자	최혜정	작성일	2021.08.03
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	데이터 구축 명확화, 변수 확정, 데이터 전처리 분배.		

## 2. 회의내용

회의내용	<p>1. 이슈 및 가치 재확인</p> <ul style="list-style-type: none"> <li>- 여름철 기상상태에 의해 전력 사용량 증가로 전력 공급 예비율(전력 공급량 - 전력사용량)이 감소함. 즉, 예비전력량 감소로 추가 설비 및 적절한 대응방안이 필요함.</li> <li>-&gt; 전력수요 예측 실패에 대한 비용 감축 가능.</li> <li>-&gt; 에너지 수요에 맞는 전력을 예측함으로써 예방 및 개선사항을 찾을 수 있음.</li> <li>-&gt; 미사용 소모 전력량이 줄지만, 적정 전력량을 예측하여 돌발 정전 사태를 예방할 수 있음.</li> </ul> <p>2. 가정사항 및 수정사항</p> <ul style="list-style-type: none"> <li>- 불쾌지수가 상승하면 일상의 가정용 전력소모가 상승할 것. 여름철 기온 상승으로 7~8월의 냉방 시설 사용이 증가하여 가정용 전력사용이 상승할 것으로 가정함.</li> <li>-&gt; 하루 전력량 예상에서는 변수가 너무 많아서 가정사항에 맞는 가정용전력으로 범위를 좁힘.</li> <li>-&gt; 여름철 기상 상태의 정의 및 조건이 필요 : 최고기온, 최저기온, 평균기온, 강수량, 평균상대 습도, 합계일사량, 불쾌지수 등으로 조건 규정.</li> </ul> <p>3. 파생변수 설정과 데이터 전처리</p> <ul style="list-style-type: none"> <li>- 기상데이터를 이용한 불쾌지수, 연평균 대비 전력 사용량을 각각의 파생변수로 결정.</li> <li>- 수집한 데이터에서 필요한 독립변수들을 가져오고 파생변수를 생성하여 데이터셋을 구축할 것.</li> </ul>
	<p>피드백</p> <ul style="list-style-type: none"> <li>- 전력 수요에 대한 변수들이 너무 광범위하기 때문에 주제를 더 구체화 할 필요가 있음.</li> <li>- 평균 전력량에 대비하여 얼마 초과 하였는지, 여름철 기상 데이터가 어떤 영향을 주는지에 생각해 보면 좋을것.</li> <li>- 분산분석(예시 - 기상 상태 범주화 후 각 범주에 따라 전력량 사용 평균이 같은지 다른지 분석 하는 것)이 적절할 것으로 보임.</li> </ul>
	<p>결정사항</p> <ul style="list-style-type: none"> <li>- 필요한 데이터를 기상데이터, 전력데이터 두 명씩 나눠서 데이터를 전처리하고 변수들의 범주화 설정에 대한 설정 기준을 마련해 볼 것.</li> </ul>
	<p>향후일정</p> <p>5차회의 : 2021.08.04, 601강의실</p>
	<p>특이사항</p>

## 3. 진행일지

작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	1) 최대전력 데이터와 점유율 결합하여 가정용전력데이터 파생변수 생성.	데이터 전처리	판다스를 이용한 데이터셋 구축 복습.
임희진	1) 기상 데이터 필요 데이터 확보 2) 16 ~ 20년 7~8월 전국 데이터 전처리 3) 각 독립변수 카테고리 범주화 후 csv 파일로 저장 4) 기획안 수정 및 피피티 초안 작성	데이터 전처리 및 기획	
배송이	1) 16 ~ 20년 7~8월 기상데이터 전처리	데이터 전처리	
우동주	1) 전력 데이터 파생 변수 생성 (연평균대비 최대전력/연평균대비가정용 최대전력)	데이터 전처리	데이터프레임 슬라이싱, 연산, 통합

# 회의록

## 1. 회의개요

일시	2021.08.04 14:00 ~ 18:00	장소	601강의실
작성자		작성일	2021.08.04
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	데이터의 종속변수에 대한 회의.		

## 2. 회의내용

회의내용	<p>1. 데이터 전처리</p> <ul style="list-style-type: none"> <li>- 여름철을 포함한 2016 ~ 2020년 기온자료와, 소비전력의 자료 통합.</li> </ul> <p>2. 변수 확인</p> <ul style="list-style-type: none"> <li>- 독립변수 : 기온, 강수량, 지면온도, 습도, 요일</li> <li>- 종속변수 : 가정용 최대전력 증감율.</li> <li>-&gt; 기온 : 기온 상승이 전력 증가에 영향을 미침.</li> <li>-&gt; 요일 : 주말과 평일에 따라 전력증가율이 다름.</li> </ul> <p>3. 변수 검토</p> <ul style="list-style-type: none"> <li>- 2016 ~ 2020년 기온변화와 전력소비량 변화의 상관관계 확인.</li> <li>- 요일을 독립변수로 설정하여 평일과 주말의 전력소비량 상관관계 확인.</li> <li>- 회귀모형에 샘플 데이터로 예측 정확도 확인</li> </ul> <p>4. 데이터 독립변수 관측치 변경</p> <ul style="list-style-type: none"> <li>- 기존: 일별 기온, 강수량, 지면온도, 습도, 불쾌지수 범주화</li> <li>- 변경: <ul style="list-style-type: none"> <li>1) 전년도 동일 기온, 강수량, 지면온도, 습도의 차이 -&gt; 회계분석 결과 정확도가 좋지 않음.</li> <li>2) 전년도 동일 기온, 강수량, 지면온도, 습도의 차이별 평균으로 나눈 후 퍼센트화</li> </ul> </li> <li>-&gt; 데이터셋 구축 전</li> </ul>
	<p><b>피드백</b></p> <ul style="list-style-type: none"> <li>- 평일 기온의 변화와 전력 소비량의 변화를 가지고 종속변수를 정해보기.</li> <li>- 기온에 따른 전력 소비량의 변화에 영향을 주는 요인을 독립변수로 설정 가능.</li> </ul>
	<p><b>결정사항</b></p> <ul style="list-style-type: none"> <li>- 요일과 같이 가정용 전력 사용의 변화에 대한 종속변수 더 탐색.</li> </ul>
	<p><b>향후일정</b></p> <p>6차회의 : 2021.08.05, 이대역그래프라운지</p>
	<p><b>특이사항</b></p>

## 3. 진행일지

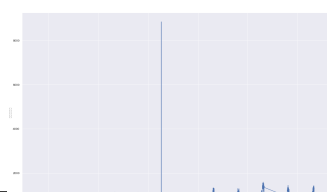
작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	1) 전력 소비량의 변화에 대한 탐색	데이터 탐색	
임희진	1) 기온 데이터 전처리 및 전체 데이터 총합 2) 독립 변수 및 종속 변수 변경 인사이트 설정 3) 각 분석모형 별 적절한 데이터 셋 구축	데이터 전처리	pandas, datetime, 모듈 전반적 사용
배송이	1) 종속변수 구체화를 위한 데이터 재탐색 2) 독립변수 데이터 수정, 상관관계 확인 후 변수 추가	데이터 전처리	
우동주	1) 연도별 가정용 전력 소비량 컬럼 생성	데이터 전처리	데이터프레임 인덱싱, for문 사용

# 회의록

## 1. 회의개요

일시	2021.08.05 14:00 ~ 18:00	장소	그래이프라운지
작성자	최혜정	작성일	2021.08.05
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	샘플링된 데이터셋으로 데이터 상관관계 확인.		

## 2. 회의내용

회의내용	<p>1. 데이터셋 샘플링. 전처리한 독립변수와 종속변수를 합쳐서 2015년 7,8월부터 2020년 7,8월 5년간의 데이터를 이용하여 데이터셋 샘플링. - 샘플링된 데이터로 상관관계 확인</p> <p>2. 데이터 추가수집. 데이터 관측치가 부족한것 같아 5년치 데이터를 추가하여 2010년부터 2020년 7,8월의 데이터를 이용하여 데이터셋을 구축하기로 함. - 기상데이터 및 전력데이터 2010~2014년 데이터 추가수집과 전처리</p> <p>3. 데이터 이상치 처리 2010~2014년 추가로 수집하여 만들어진 데이터셋 확인 후 전력데이터 시각화</p>  <p>&gt; 전력데이터 2014년 7월 11일 이상치 발견.</p>
	<p>- 종속변수인 전력데이터를 더 자세히 분석하여 기온변화에 어떤것들이 영향을 미쳤는지에 대하여 더 찾아보기. 즉, 독립변수를 더 개발하기. - Day를 레이블인코딩으로 변경하기. - 추가적인 데이터 수집하면 더 좋을것.</p>
결정사항	<p>- 전력데이터 이상치 수정하기. - 여름철 전력사용량에 영향을 미친 이벤트 찾아보기</p>
향후일정	7차회의 : 2021.08.06, 이대역그래이프라운지
특이사항	

## 3. 진행일지

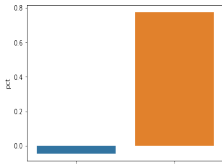
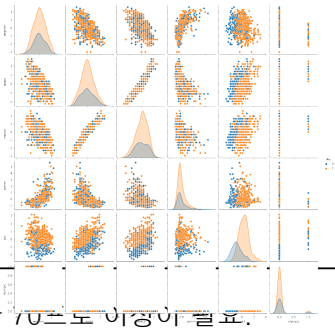
작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	1) 전력데이터 2009~2014년 자료 수집 2) 가정전력 계산.	데이터 전처리	for문 복습, 판다스이용.
임희진	1) 2010~ 2013데이터 수집 2) 데이터 상관관계 확인 3) 독립변수 구축 샘플링 4) 회귀분석 모델에 맞는 데이터 셋 구축 샘플링	데이터 전처리 및 활용방안 탐색	
배송이	1) 2009년, 2013~2014년 기상데이터 수집 2) 기존 데이터와 합쳐서 전처리	데이터 전처리 및 탐색	
우동주	1) 2010~2015년 가정용 점유율 수집 2) 가정용 전력 증감을 파생 변수 생성	데이터 전처리	시각화를 통한 이상치 탐색

# 회의록

## 1. 회의개요

일시	2021.08.06 14:00 ~ 18:00	장소	그레이프라운지
작성자	최혜정	작성일	2021.08.06
참석자	팀장 임희진 / 팀원 우동주, 배송이, 최혜정		
안건	데이터셋 최종 검토.		

## 2. 회의내용

회의내용	<p>1. 데이터셋 구축 후 특이사항 발견.</p> <ul style="list-style-type: none"> <li>- 전예비전력 20%를 고려하여 전년동기대비 20% 이상 증가한 날짜 확인.</li> <li>- 확인한 년도 및 날짜에 특별한 일이 있었는지 검색.</li> </ul> <p>2. 추가 변수 발견.</p> <ul style="list-style-type: none"> <li>- 전년동기대비 20% 이상 증가년도 2012년, 2016년, 2018년 에 영향을 미치는요인에는 기록적인 폭염을 동반하여 특별한 일이 있었는지 검색.</li> <li>- 2012년과 2016년 에는 하계올림픽이 있었던 것으로 발견됨.</li> </ul>		
	 <p>올림픽이 없을때(0)와 있을때(1)의 증감을 비교 -&gt; 올림픽을 변수로 추가설정.</p> <p>3. 최종 데이터셋 시각화</p> <ul style="list-style-type: none"> <li>- 데이터들의 상관관계를 시각적으로 확인하고 결과 분석.</li> </ul> 		
피드백	<p>R2 ...</p> <p>... 하지만 70프로 이상이 필요.</p> <ul style="list-style-type: none"> <li>- 탐색적 부분 결과를 정리하기.</li> </ul>		
결정사항			
향후일정	발표 2021.08.09		
특이사항			

## 3. 진행일지

작성자	계획	과정	기록(복습, 배운내용, 사용기술 등)
최혜정	전처리과정 정리.	마무리	for문 이용하여 데이터셋 구축 간단화
임희진	1) 데이터 최종 정리 2) 결과 report 작성 3) 결과 ppt 작성	마무리	
배송이	전처리과정 정리.	마무리	피벗테이블, 데이터프레임 연산, 결합
우동주	전처리과정 정리.	마무리	for문, 피벗테이블 이용에 대한공부

프로젝트 단계	참여자	시작일	작업일수	종료일	진행현황	진행일수	비고
주제, 데이터 탐색	임희진	2021-07-30	1	2021-07-31	100%	1	
	배송이	2021-07-30	1	2021-07-31	100%	1	
	우동주	2021-07-30	1	2021-07-31	100%	1	
	최혜정	2021-07-30	1	2021-07-31	100%	1	
주제 선정, 방향 설정	임희진	2021-07-31	1	2021-08-01	100%	1	
	배송이	2021-07-31	1	2021-08-01	100%	1	
	우동주	2021-07-31	1	2021-08-01	100%	1	
	최혜정	2021-07-31	1	2021-08-01	100%	1	
데이터수집	임희진	2021-07-31	2	2021-08-02	100%	2	
	배송이	2021-07-31	2	2021-08-02	100%	2	
	우동주	2021-07-31	2	2021-08-02	100%	2	
	최혜정	2021-07-31	2	2021-08-02	100%	2	
인사이드설정	임희진	2021-07-31	2	2021-08-02	100%	2	
	배송이	2021-07-31	2	2021-08-02	100%	2	
	우동주	2021-07-31	2	2021-08-02	100%	2	
	최혜정	2021-07-31	2	2021-08-02	100%	2	
데이터전처리	임희진	2021-08-01	4	2021-08-05	100%	4	
	배송이	2021-08-01	4	2021-08-05	100%	4	
	우동주	2021-08-01	4	2021-08-05	100%	4	
	최혜정	2021-08-01	4	2021-08-05	100%	4	
활용방안 코딩진행	임희진	2021-08-03	2	2021-08-05	100%	2	
	배송이	2021-08-03	2	2021-08-05	100%	2	
	우동주	2021-08-03	2	2021-08-05	100%	2	
	최혜정	2021-08-03	2	2021-08-05	100%	2	
마무리 진행	임희진	2021-08-05	1	2021-08-06	100%	1	
	배송이	2021-08-05	1	2021-08-06	100%	1	
	우동주	2021-08-05	1	2021-08-06	100%	1	
	최혜정	2021-08-05	1	2021-08-06	100%	1	
완성 및 발표	임희진	2021-08-06	0	2021-08-06	100%	0	
	배송이	2021-08-06	0	2021-08-06	100%	0	
	우동주	2021-08-06	0	2021-08-06	100%	0	
	최혜정	2021-08-06	0	2021-08-06	100%	0	



