

Wavve 경력 기술서

Table of Contents

데이터개발팀	2
` 최신 AI 기술 연구로 지속적인 학습과 개선 도모	3
` 대화형 AI 챗봇 프로젝트로 사용자 경험 향상.....	4
` 클라우드 이전으로 데이터 마트 생성 시간 90% 이상 감소.....	6
` 맞춤형 데이터 파이프라인 최적화로 비용 30% 이상 절감	7
` 멘토링 및 지식 공유를 통한 데이터 플랫폼 환경 교육.....	9
` 대시보드 재설계를 통한 연간 \$4,200 절감 및 새로고침 시간 단축.....	10
` #data_communication Slack 채널 개설로 데이터 지식 공유 효율화	12
경영전략부	14
` 데이터 마트 환경 재구축을 통한 프로세스 개선 및 효율성 증대	15
` 데이터 기반 전략 수립과 분석 문화 확산	16
데이터마케팅팀.....	17
` 통계적 통찰력 도출 및 대시보드 디자인을 통한 운영 효율성 향상	18
` CRM 을 통한 성과 지표 설계 및 사용자 세분화 분석	19

데이터개발팀

wavve

“ 최신 AI 기술 연구로 지속적인 학습과 개선 도모

From: 2023-04 ~ Current

기술 분야에서의 최신 트렌드와 업데이트를 주시적으로 따르고 중요하게 생각하고 있습니다. 특히, 2022 년 11 월부터 끊임없이 논의되었던 ChatGPT 서비스에 대한 큰 관심을 가지고 있어, 제 팀원들과 기술 부서 동료들과 함께 새로운 모델과 기술 업데이트에 대한 관련 기사를 공유하고 있습니다. AI 기술의 인기가 높아지면서 기업에서도 AI 모델과 관련된 연구와 개발이 필요하다고 판단하게 되었고, 이에 따라 2023 년 4 월부터 AI 연구 부문을 주도적으로 맡게 되었습니다.

이 포지션에서, 저는 주요 클라우드 제공 업체인 AWS, Azure 및 Google 과 긴밀하게 협력하면서 비즈니스 미팅을 진행하여 최신 모델과 시장 동향에 대한 정보를 빠르게 습득하고 있습니다. 각 회사의 솔루션 아키텍트와의 비즈니스 이메일을 통해 새로운 기술의 사용 사례와 기능에 대해 명확하게 이해하고자 적극적으로 의사 소통에 참여하고 있습니다.

OTT 회사에서 콘텐츠를 다루는 일원으로서, 텍스트, 이미지, 비디오 모델을 주로 다루었습니다. 클라우드 제공 업체 외에도 출시된 서비스들을 테스트하면서 현재 기술의 수준을 확인하고, Wavve 서비스에 적용할 수 있는 새로운 아이디어를 모색했습니다. 상반기 동안, 팀원들과 함께 AI 모델을 사용하면서 테스트하고, 그 결과를 정리하여 기술 행사에서 공유하고 기술 블로그에 기고문을 작성하는 기회 또한 얻게 되었습니다. 프레젠테이션과 글 작성을 통해 새로운 기술에 대한 이해를 높이고, 내용을 정확히 이해했는지 제 자신을 전문적으로 평가하고 확인 할 수 있었습니다.

생성형 AI, '밥그릇 도둑'인가? '생산적 도구'인가?

이 경험을 통해, 다양한 프로그래밍 작업을 수행하며 지식을 공유하고 다른 사람들이 새로운 기술을 탐구하도록 격려하는 것이 혁신적인 작업 환경을 조성하는 데 중요하다는 것을 이해하게 되었습니다. 이를 기반으로, 최신 기술 트렌드를 계속 연구하고, 이를 통해 팀의 성장과 개선을 이끌어내는 데 노력하고 있습니다. 이러한 노력을 인정받아 현재 Wavve 대화형 검색 프로젝트를 주도적으로 진행하게 되었습니다.

`` 대화형 AI 챗봇 프로젝트로 사용자 경험 향상

From: 2023-06 ~ Current

AI 리서치를 진행하면서 사용자들이 원하는 정보에 빠르게 접근하고, 편리하고 재미있는 정보 검색 경험을 제공하는 기회를 발견했습니다. 이를 위해 대화형 AI 챗봇 프로젝트를 주도하게 되었으며, 주요 목표는 대화형 검색을 통해 맥락 시나리오를 분석하고 사용자 요구 사항을 파악하는 것이었습니다. 이 프로젝트에서 제가 담당한 역할은 기획 보고서 작성, 데이터 개인 정보 보호 및 보안 보장, 서비스 디자인, 백엔드 및 프론트엔드 리소스 관리, 모델 성능 평가를 위한 프롬프트 엔지니어링, 챗봇 시스템 통합을 위한 메타데이터 구축 등 다양한 영역을 포괄하였습니다.

현재 웨이브에서 만개가 넘는 다양한 장르의 VOD 서비스를 제공하고 있습니다. 많은 콘텐츠를 보유하고 있음에도 불구하고 홈 화면에 제한된 콘텐츠를 표시해야 하는 제약이 부분이 있어 효율적으로 콘텐츠를 노출하기 위한 데이터 분석은 필수입니다. 이로 인해 이용자들에게 편리한 검색 경험을 어떻게 제공할 수 있을지에 대한 고민이 시작되었고, 대화형 검색 채널을 개발함으로써 실제 사용자 요구를 자연어로 수집하고, 수많은 콘텐츠 중에서도 가장 관련성 높은 프로그램을 검색하는 기능을 구현하고자 했습니다. 챗봇이 사용할 데이터를 인덱싱하고 관리하기 위해 클라우드 검색 서비스인 Cognitive Search 를 사용했으며, Azure 에서 제공되는 Prompt Flow 와 Cognitive Search 를 연결하여 Prompt Engineering 작업을 개발하고, Bulk Test 기능을 통해 여러 질문 결과를 신속하게 평가하여 Prompt Engineering 의 개발 속도를 향상시켰습니다.



Prompt Engineering Sample

--- You session start from here. ---

멀티버스가 뭐야?

멀티버스는 '다중 우주'를 의미하는 용어로, 우리가 살고 있는 우주 외에도 무수히 많은 다른 우주들이 존재한다는 이론을 가리킵니다. 이는 과학, 철학, 그리고 특히 공상과학(Sci-Fi) 작품에서 자주 다루어지는 주제입니다. 멀티버스에 대한 이해를 돕기 위해 Wavve에서 찾을 수 있는 관련 콘텐츠를 추천드립니다: 1. '멀티버스 프로젝트' 2. '(더빙) 닥터 스트레인지: 대혼돈의 멀티버스' 3. '(자막) 닥터 스트레인지: 대혼돈의 멀티버스' 4. '[확장판] 에브리씽 에브리웨어 올 앳 원스+'. 이 목록은 Wavve(웨이브) 검색 결과에서 제공된 것입니다.

Total tokens for generating this: 2810 tokens, time spent: 15.63 sec

멀티버스 프로젝트는 무슨 내용이야?

'멀티버스 프로젝트'는 의문의 장소에서 감금된 채 깨어난 '제인'의 이야기를 담고 있습니다. 제인은 그곳을 탈출하려 하지만 어느 순간 다시 정신을 잃고 깨어나는 것을 반복합니다. 연이은 시도 끝에 가까스로 탈출에 성공하지만, 거리에는 인가척이 느껴지지 않습니다. 가끔 어디선가 신원미상의 목소리만 들릴 뿐입니다. 차고에 버려져 있던 우편물을 단서로 그녀는 무작정 '위트니스 연구 센터'로 향하고 그곳에서 놀라운 광경을 목격합니다. 이 작품은 Wavve(웨이브)에서 확인하실 수 있습니다.

Total tokens for generating this: 3054 tokens, time spent: 15.08 sec

다른거 또 뭐 없어?

멀티버스에 관한 콘텐츠를 찾으시는군요! Wavve에서 찾을 수 있는 멀티버스 관련 콘텐츠는 다음과 같습니다: 1. '멀티버스 프로젝트' 2. '(더빙) 닥터 스트레인지: 대혼돈의 멀티버스' 3. '(자막) 닥터 스트레인지: 대혼돈의 멀티버스' 4. '[확장판] 에브리씽 에브리웨어 올 앳 원스+' 5. '(자막) 스파이더 맨: 어크로스 더 유니버스' 이 목록은 Wavve(웨이브) 검색 결과에서 제공된 것입니다.

Total tokens for generating this: 3637 tokens, time spent: 13.83 sec

① Only default variants will be used for chat, if you want to test variants please try bulk test. For chatbot and test app bot, it will only show the chat output.

현재 해당 프로젝트는 기존 로드맵에 없던 계획으로, 최소한의 리소스로 프로젝트를 진행하고 있지만, 팀원 모두 AI 분야에서 비교적 최신이라 할 수 있는 프롬프트 엔지니어링에 대해 깊이 공부할 수 있는 좋은 기회로 여기며 작업에 동참하고 있습니다. 저 또한 프로젝트 관리자로서 데이터 스키마 설계부터 데이터 분석까지 데이터 흐름을 종합적으로 이해할 수 있는 기회를 얻게 되어, 해당 경험을 통해 데이터 생태계에 대한 지식이 한층 더 깊어질 것이라고 생각합니다.

`` 클라우드 이전으로 데이터 마트 생성 시간 90% 이상 감소

From: 2022-04 ~ 2022-07

2022년 4월 데이터응용개발팀에 합류할 당시 Wavve 플랫폼기술본부에서는 Google Cloud Platform을 도입하여 환경을 구축하고 있었습니다. 매체전략부에서 데이터 마트 생성 쿼리 성능 향상 작업을 담당하면서 얻은 데이터에 대한 이해와 테이블 구조 지식을 활용하여 데이터응용개발팀에서의 첫 작업으로 데이터 처리시간이 긴 기존 온프레미스 Airflow 데이터 마트 환경을 구글 클라우드의 BigQuery와 Cloud Composer 환경으로 파이프라인 및 테이블 이전 작업을 계획 & 진행하였습니다.

 DM데이터 생성시간 비교

 소유자: 노윤정 ...
7월 21, 2022

- WAVD-4553: [IM] DM 데이터 생성시간 비교 종료
- GCP : 346초 소요
- Azure : 7286초 소요
- 상세 비교

DAGs	TASKs	작업소요시간 (wavvedp-dev 환경기준)	기존 Azure 작업소요시간
AZ_DM_1_OTHERS	payment_step	12	39
	retry_success	7	75
	sales_movie_job	6	170
	join_job	5	31
	demographics	9	

파이프라인 및 테이블 이전 작업을 설계 & 모니터링을 하면서, BigQuery의 뷰 및 테이블 만료 날짜 기능을 전략적으로 활용하여 관리되는 데이터 테이블 수를 22개에서 13개로 줄일 수 있었습니다. 이렇게 간소화된 데이터 관리 환경으로 인해, 데이터 마트 생성 시간을 대략 90% 이상 감소하여 데이터 효율성이 크게 향상되었습니다.

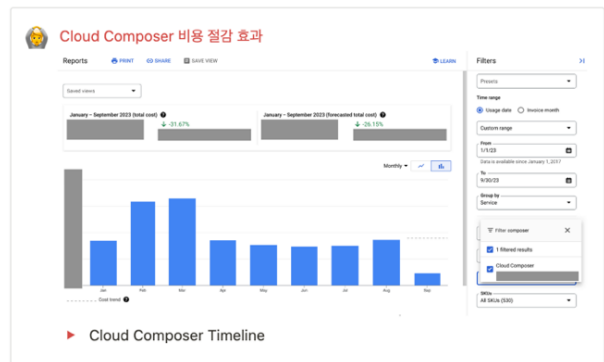
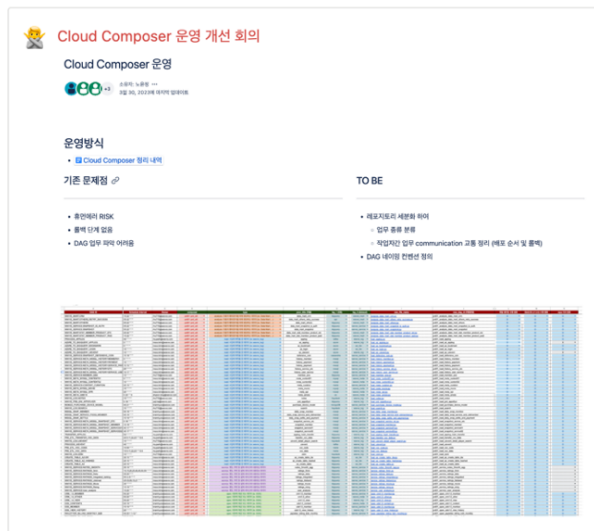
기술본부의 인원으로서 담당하게 되는 첫 작업이어서 긴장도 많이 했습니다만 그동안 쌓아왔던 데이터 도메인 지식과 테이블 구조를 꾸준히 기록하면서 작업해왔기에 큰 어려움 없이 데이터 마트 파이프라인 생성 환경을 클라우드로 이전할 수 있었습니다. 또한 혼자만 알고 있었던 데이터 마트에 대한 지식을 팀원들과 공유함으로써 정보에 대한 지식의 부담감을 내려 놓을 수 있었던 좋은 작업 기회였습니다.

`` 맞춤형 데이터 파이프라인 최적화로 비용 30% 이상 절감

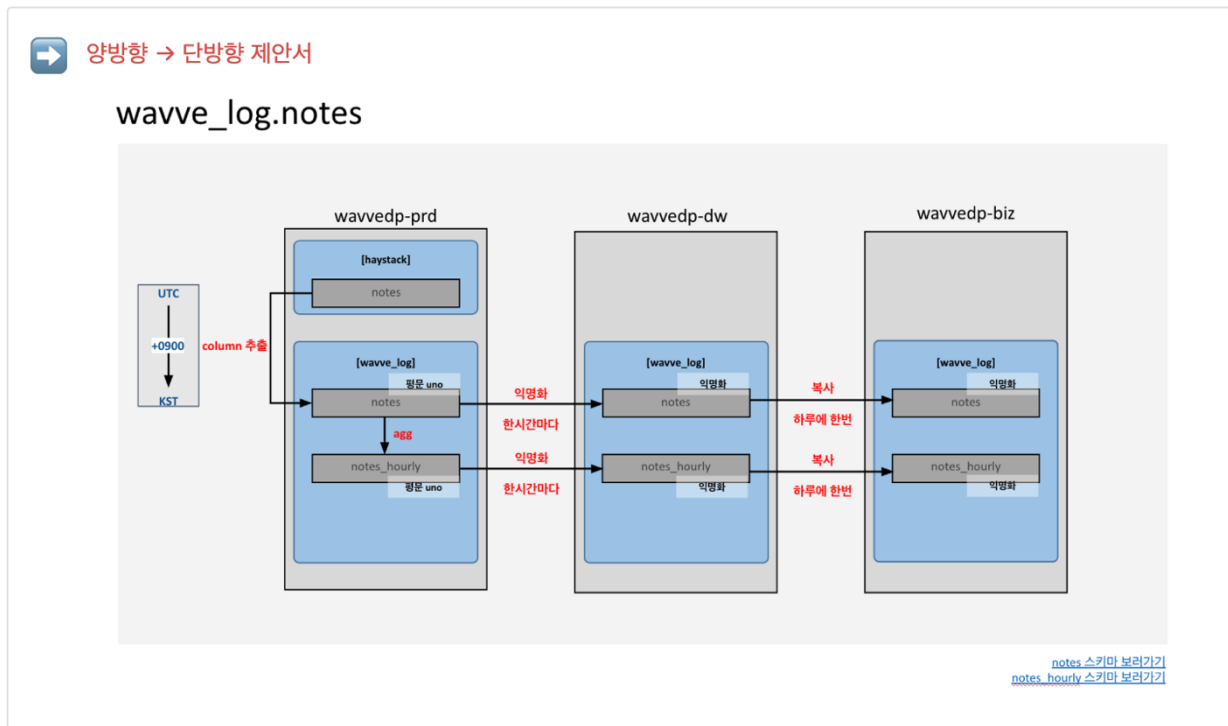
From: 2022-04 ~ 2023-04

현재 Wavve 데이터 거버넌스는 데이터개발팀이 주도적으로 담당하고 있습니다. 데이터개발팀이 관리하는 GCP 프로젝트는 상용 & 개발 프로젝트 모두 포함하여 총 5개입니다. 프로젝트마다 데이터 보존 및 특정 요구 사항이 다르며, 이에 따른 비용 및 자원 최적화를 위한 파이프라인 개선이 필요하다고 판단하였습니다. 또한, 정보 보안팀과 협력하여 RBAC 관리를 통해 보안을 강화하고, 양방향 파이프라인 문제를 단방향으로 재구성할 필요성을 인식하였습니다.

각 프로젝트 특성에 맞게 Cloud Composer 환경을 구성하여 비용 최적화에 집중함과 동시에 Composer에 배포되는 Airflow DAG에 대한 정책과 네이밍 컨벤션을 수립하여 독립적인 데이터 파이프라인을 통해 작업자간 배포 영향도를 최소화 할 수 있도록 진행했습니다. 팀원들과 회의를 통해 데이터 파이프라인 작업 종류를 7개의 분류로 나눠 담당자 부재시에도 빠르게 문제가 있는 파이프라인을 확인할 수 있도록 tag 정책과 prefix 규칙을 정하여 배포를 하도록 기준을 잡고 공지했습니다. 작업당시 데이터 파이프라인을 작업하는 팀 인원은 7명으로 작업 초반 당시 미팅 시간을 조율하고 작업을 나누는 기준을 잡기 어려웠지만 최소한의 정보를 입력해도 네이밍 컨벤션을 자동으로 반영할 수 있도록 엑셀 파일에 함수를 걸어 제공하고 필요시 멤버와의 1:1 미팅을 통해 최적화 작업을 이어 나갔습니다. 이를 통해, 대략적으로 30% 이상의 비용을 절감할 수 있었습니다.



또한, 현재 데이터개발팀이 담당하고 있는 GCP 프로젝트는 개인정보 데이터 포함 유무로도 구분되어 관리되고 있습니다. 정보 보안팀과의 협력으로 프로젝트의 접근을 RBAC 관리로 강화하였으며, 기존에 구축되어 있던 양방향 파이프라인 문제를 단방향으로 재구성하여 보안성을 높였습니다. 프로젝트 데이터 파이프라인 구축 초기 당시 양방향 파이프라인의 문제점을 빠르게 인식 후 단방향으로 수정 완료한 덕분에 23년 4월 기존 가명화 처리되고 있었던 개인정보 데이터를 익명화 솔루션인 Data Genor를 도입하여 460TB가 넘는 시정 데이터를 큰 이슈없이 모두 익명화처리 하여 보안성을 강화할 수 있었습니다.



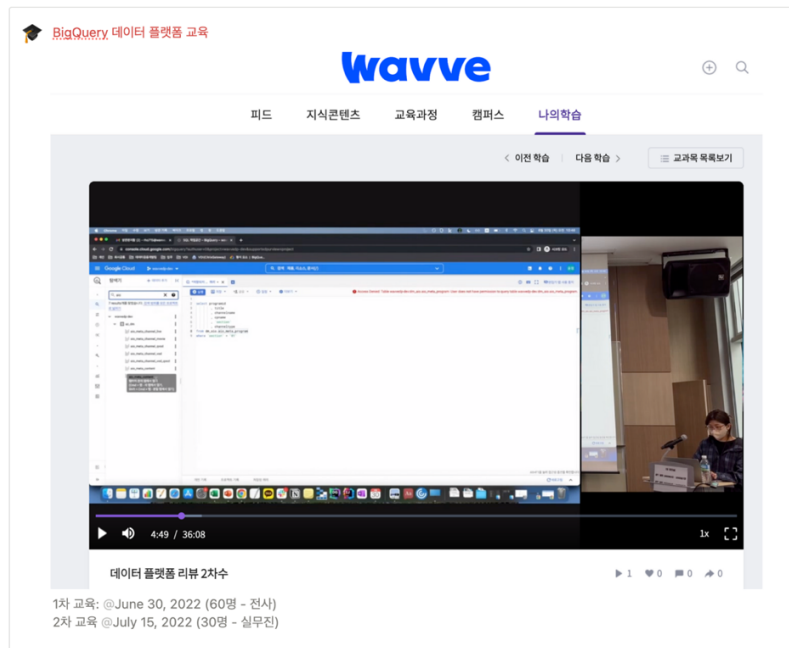
데이터 파이프라인 최적화 작업을 통해 클라우드 환경에서 비용 효율을 확인했으며, 이 경험을 통해 클라우드 기반 생태계의 운영 효율성의 중요성을 인지했습니다. 더불어 최적화 작업은 데이터 파이프라인을 종합적으로 이해하고 전체 데이터 웨어하우스 이해도를 향상시켰습니다. 이러한 과정을 통해 성능 최적화와 강력한 보안 조치의 중요성을 높게 인식하게 되었습니다.

“ 멘토링 및 지식 공유를 통한 데이터 플랫폼 환경 교육

From: 2022-04 ~ 2022-07

데이터 분석을 적극적으로 사용하려는 다른 부서들의 필요성을 인지하였습니다. 회사의 데이터 종류와 도메인 지식을 공유하고, 개발팀 & 비개발팀 모두에게 BigQuery 환경 내에서의 데이터 활용 방법에 대한 설명과 SQL 쿼리 사용에 대한 교육을 전달하려는 목표를 설정하였습니다.

전사와 실무진 대상으로 1, 2 차 교육을 나눠서 진행했습니다. 각 교육 대상에 맞는 간결한 프레젠테이션을 제작하고 실습 문제를 제공하여 데이터 분석을 적극적으로 관리할 수 있도록 지원하였습니다. 이를 통해 협력과 데이터 학습 중심의 팀 환경을 조성할 수 있도록 크게 노력하였습니다.



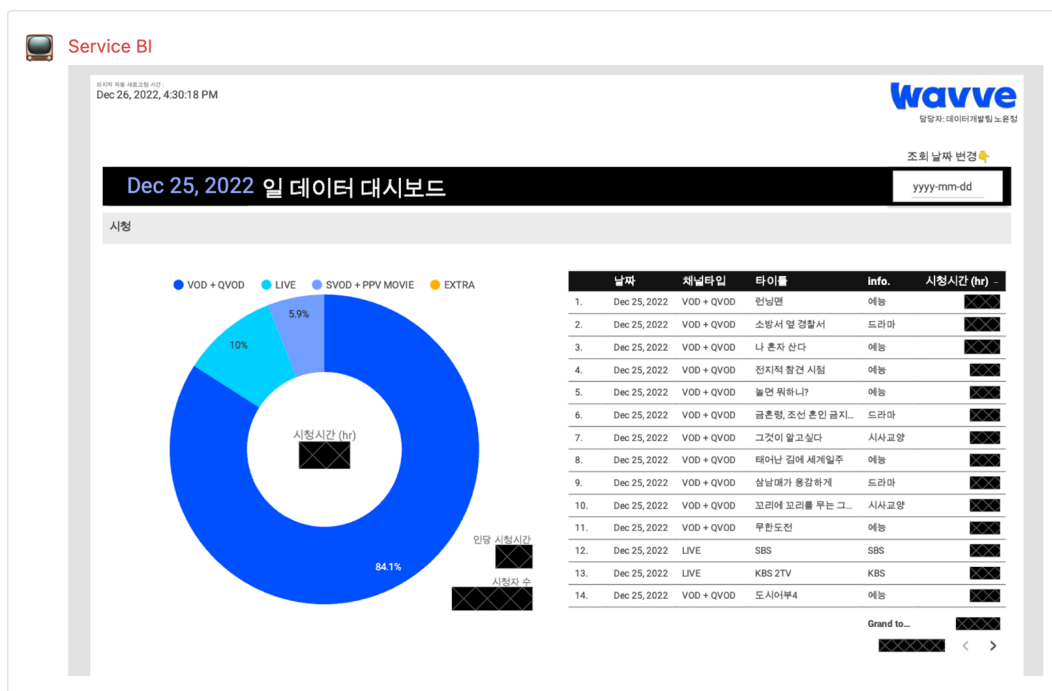
적절한 교육과 지식 공유를 통해 팀의 효율성과 협력성을 향상시킬 수 있다는 것을 알게 되었습니다. 현재 비개발팀 부서의 쿼리 사용 인원이 약 3배정도 증가하였으며 다양한 분석 자료들이 제작되고 전사적으로 공유가 진행되고 있습니다. 현재 추가적인 교육은 진행하고 있지는 않지만 #data_communication의 오픈 슬랙 채널을 통해 지속적인 데이터 Q&A 세션을 운영함으로써 회사의 전체적인 데이터 활용 능력을 향상시키는 데 기여하고 있습니다.

`` 대시보드 재설계를 통한 연간 \$4,200 절감 및 새로고침 시간 단축

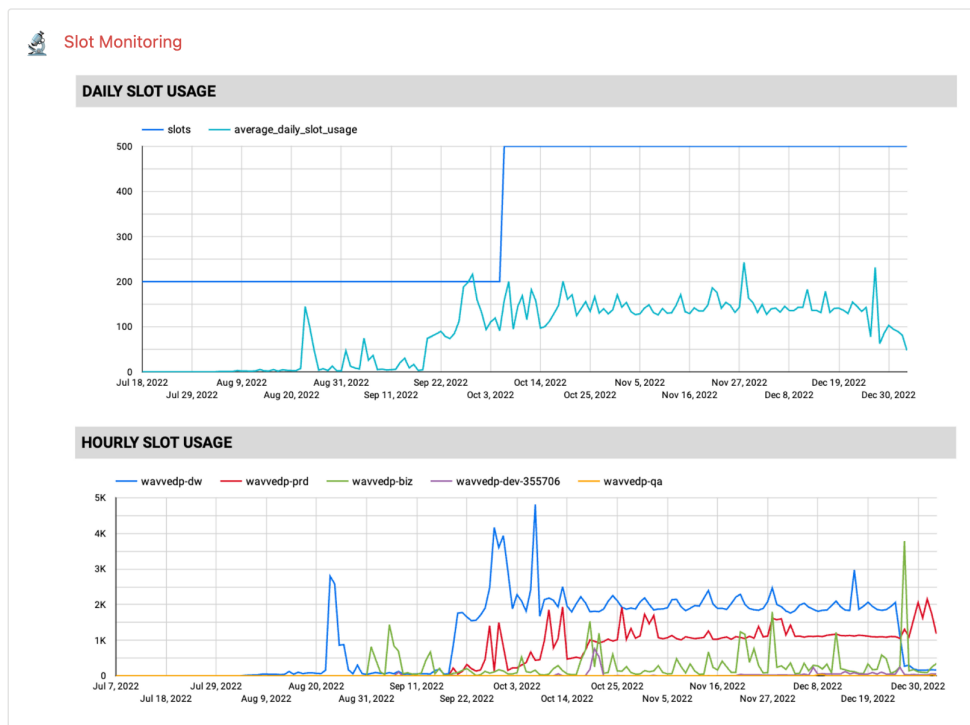
From: 2022-04 ~ 2022-12

BI 도구인 Tableau 사용에 따른 연간 라이선스 비용과 보고서 새로고침 시간이 길어지는 문제를 인식하였습니다. 이를 해결하기 위해 더 효율적인 도구를 선택할 필요성을 느꼈고, Google Cloud Platform 도입계획에 따라 Data Studio로의 대시보드 전환을 계획하였습니다.

우선적으로 매주 월요일 오전 8시에 임원회의에서 사용되는 Service BI Tableau 리포트를 Data Studio로 재구현했습니다. 기존 Tableau 리포트에서는 데이터를 불러오는데 많은 제약이 있어 많은 정보를 하나의 리포트 안에 담아내지 못하는 어려움이 있었습니다만 Data Studio로 환경을 고도화하게 되면서 시청시간, 유료가입자 추이는 물론 데이터마케팅 소속 당시 개발하였던 First View의 지표 등 좀 더 다양한 사용자 행동에 대한 데이터를 포함하여 Wavve에서 서비스되고 있는 프로그램에 대한 통찰력을 보다 효과적으로 제공할 수 있도록 사용자 친화적인 대시보드를 작성하였습니다.



또한 Google Cloud 환경의 장점을 살려 Cloud 환경의 데이터를 BigQuery로 저장하여 BigQuery 슬롯 사용량 및 청구 비용 모니터링 대시보드를 제작하여 비용 절감에도 효율적으로 사용할 수 있도록 대시보드를 제공했습니다. 오랜기간 Tableau를 관리하면서 익힌 그래프와 매개변수 경험을 통해 비록 새로운 데이터와 새로운 툴을 다루는 작업이었음에도 불구하고 빠르게 대시보드를 성공적으로 전환하여 Tableau 연간 라이선스 비용 \$4,200을 절감하며 보고서 새로고침 시간또한 1시간에서 단 10분으로 단축시킬 수 있었습니다.



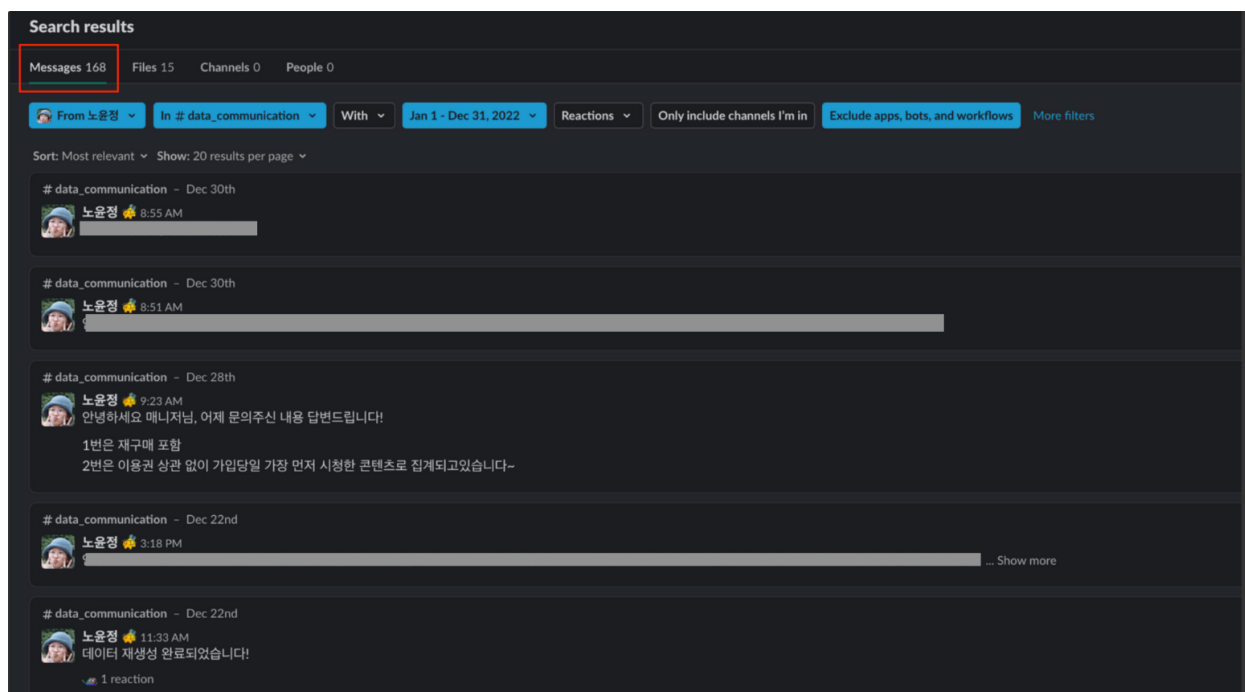
해당 경험을 통해, 적절한 도구 선택 및 전환을 통해 비용 절감과 작업 효율성 향상을 실현할 수 있는 중요성을 깨닫게 되었습니다. 또한, 클라우드 환경에서의 데이터 처리를 통해 GCP 환경에 대한 심도 있는 이해를 높일 수 있었습니다. Tableau 와 같은 BI 도구에 대한 사전 지식 덕분에 관련 지식을 신속하게 습득하고 작업할 수 있었으며, 이를 통해 다양한 시각화 경험을 쌓음으로써 새로운 도구 사용에 대한 흥미를 얻을 수 있었습니다. 또한, Tableau 를 통해 생성된 대시보드뿐만 아니라 분석에 필요한 추가 대시보드를 작성하고, 현재까지도 대시보드 성능을 원활하게 유지하기 위한 지속적인 모니터링을 담당하고 있습니다.

`` #data_communication Slack 채널 개설로 데이터 지식 공유 효율화

From: 2022-04 ~ Current

이전 매체전략팀에서의 역할로, 데이터 분석 방법론, 결과 및 통찰력을 문서화하고, 팀 구성원 간의 참고 및 지식 공유의 필요성을 인식하였습니다. 이에 따라, **#data_communication** 이라는 Slack 채널을 개설하였습니다. 데이터개발팀으로 이직하면서 해당 채널의 운영 역할 또한 자연스럽게 데이터개발팀으로 이관되었으며 해당 채널은 데이터 거버넌스 역할이 데이터개발팀으로 원활하게 자리잡는 데 중요한 역할을 하였습니다.

#data_communication Slack 채널의 운영이 데이터개발팀의 역할로 이관되면서 해당 채널을 통해 BigQuery 환경의 변경 사항/업데이트를 공유하고, 데이터 분석 및 쿼리 문의에 답변을 구하는 등의 활동을 확대하였습니다. 이를 통해, 기술 및 비기술적 이해 관계자 간의 소통 간극을 줄이고, 회사 내 모든 사람이 우리의 서비스 및 구독 시스템에 대한 더 나은 이해를 가질 수 있도록 지원하였습니다. 적극적인 소통을 강조하며 동기를 부여하였고, 2022년에는 채널에 총 771 개의 메시지 중 168 개를 작성하여, 두 번째로 활동적인 팀 구성원이 작성한 87 개의 메시지보다 거의 두 배 가까이 기여하였습니다.



데이터 이해뿐만 아니라 이러한 지식을 열정적으로 공유하여 조직에 이점을 제공하는 것을 항상 중요하게 여겼습니다. 또한, 비개발자로서 커리어를 시작하여 개발자와의 소통 방식을 익히고 적절한 커뮤니케이션 도구를 활용하여 지식 공유를 효율화하고, 팀의 데이터 이해도를 높일 수 있다는 것을 깨달았습니다. 이를 기반으로 데이터 관련 지식 공유를 계속 활성화하고, 기술 및 비기술적 이해 관계자 간의 간극을 줄이는 데 노력하며 조직 전체의 데이터 활용 역량을 향상시켜 모두가 공통의 목표를 추구하며 협업할 수 있도록 노력하고 있습니다.

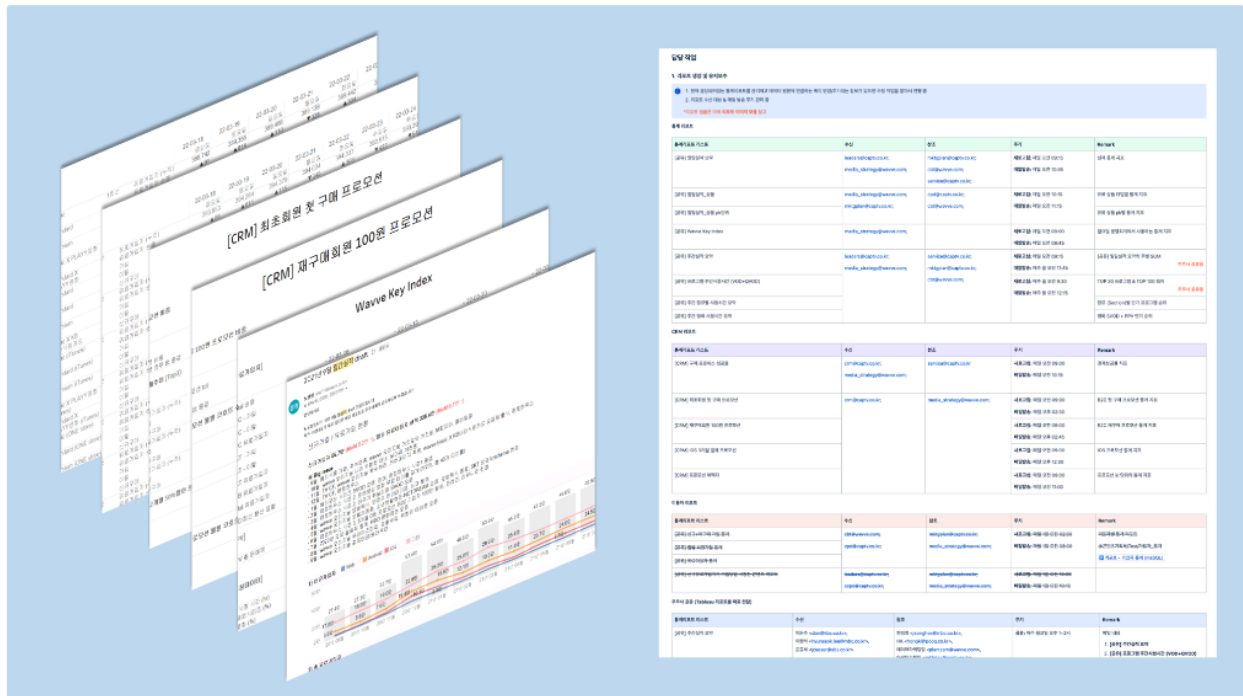
경영전략부

wavve

`` 데이터 마트 환경 재구축을 통한 프로세스 개선 및 효율성 증대

데이터 마트 담당자로 지정되면서 SQL 처리 시간 감소 및 쿼리 성능 향상을 통해 데이터 품질 개선을 목표로 인수인계를 진행하였습니다.

SQL 전문 지식을 활용하여 중복되는 서브 쿼리를 최적화하고, 임시 테이블 및 WITH 문을 활용하여 데이터 마트 쿼리를 최적화하였습니다. 이를 통해 21 개의 데이터 마트 테이블 생성에 사용되는 32 개의 쿼리 파일을 수정하였고, 직렬로 작동되던 Airflow Task 구조를 병렬화하여 SQL 처리 시간을 20% 감소시키는 데 성공했습니다. 더불어 데이터 품질 검사 및 유효성 검사 프로세스를 구현하여 데이터의 정확성과 무결성을 확보하였습니다.

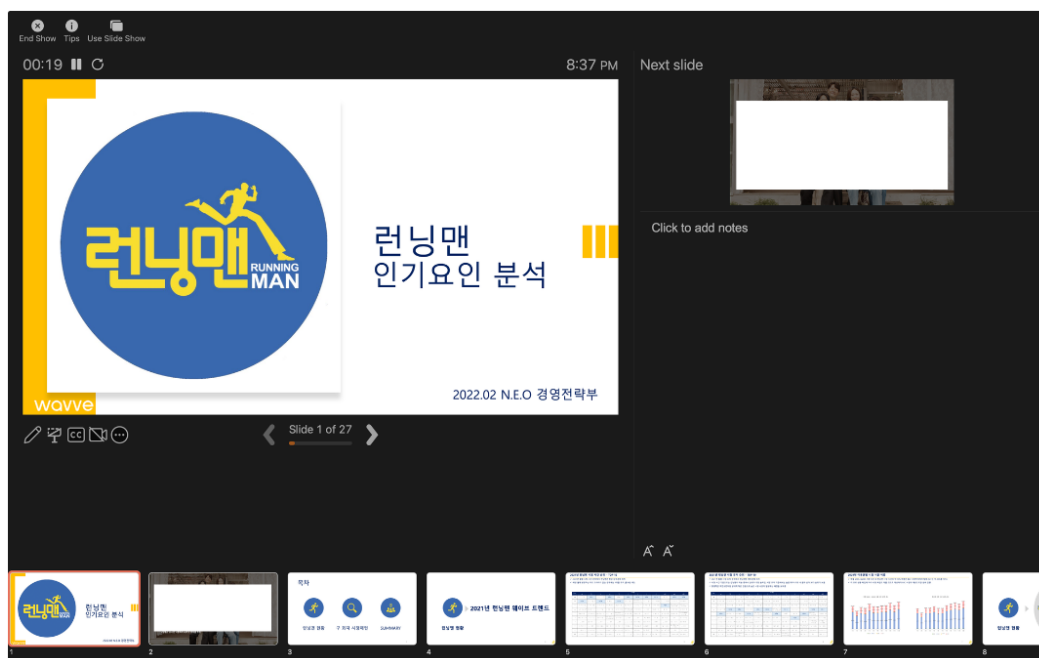


누락된 데이터를 추적하기 위한 모니터링 대시보드를 개발하고, 이상점을 식별하기 위해 Tableau 의 conditional formatting 및 매개변수와 같은 기능을 활용하여 그래프를 생성했습니다. 더불어 데이터 조회 속도의 향상으로 기존에 잘게 나누어 있던 대시보드를 효율적으로 하나로 통합하여 주간 회의 자료로 활용하며, 이로써 리더 및 고위 임원에게 실행 가능한 통찰력을 제공하였습니다.

“ 데이터 기반 전략 수립과 분석 문화 확산

데이터 기반 전략을 수립하기 위해 적극적으로 SQL 을 활용하여 사용자 행동 및 구독 데이터에서 트렌드, 패턴, 이상 현상을 식별하고, 이를 통한 의사 결정을 진행해왔습니다. 지속적으로 데이터에 대한 관심을 유도하고 회사의 구독 프로세스와 데이터 기반 통찰력을 이해하는 데 중점을 두어 사용자 참여, 구독 유지 및 콘텐츠 성능 분석을 수행했습니다.

웨이브 플랫폼에서는 오랜 기간 동안 사용자들로부터 사랑을 받는 프로그램 중 하나로 SBS 예능 프로그램 '런닝맨'이 있습니다. 웨이브 이용자를 더 깊게 이해하기 위해 '런닝맨'의 인기 요인 분석을 주도하였으며, 이를 통해 필요한 데이터를 팀원들과 협력하여 추출하고 리포트를 준비하면서 데이터 기반 통찰력을 촉진했습니다. 분석 진행 중에는 패턴 및 이상 현상을 식별하여 특정 디바이스에서 시청시간 중복 발생 패턴을 빠르게 감지하고 기술본부와 협력하여 문제를 해결했습니다.



해당 리포트는 다른 부서가 추가 분석을 진행하는 데 도움이 되었으며, 분석 과정에서 발견한 시청시간 오류에 대한 해결 방안을 찾아가는 과정에서 데이터 파이프라인에 대한 궁금증과 해결 방법을 탐구하고자 전략본부에서 기술본부로 조직 이동을 결정하게 되었습니다.

데이터마케팅팀

wavve

“ 통계적 통찰력 도출 및 대시보드 디자인을 통한 운영 효율성 향상

데이터마케팅팀 입사 후, 데이터 추출과 제공 과정에서 다양한 어려움이 발생했습니다. 회사의 데이터는 다양한 소스에서 수집되며, 형식과 구조가 제각각이었고, 스키마 문서의 부재로 데이터 이해와 처리에 어려움이 있었습니다. 또한, 부서 간 데이터 관련 용어와 요구사항이 혼재되어 혼란이 발생하며, 요청 및 제공 과정이 더 시간이 소요되는 경우가 있었습니다.

데이터 구조를 이해하기 위해 데이터 딕셔너리를 작성하고, 구글 시트를 활용하여 데이터의 의미와 구조를 문서화 했습니다. 이로써 부서 간 커뮤니케이션을 효율화하고 일관성을 유지했습니다. 또한, 정기 회의를 통해 부서 간 요구사항을 조율하고 표준 용어 및 프로세스를 정립하는 작업을 진행했으며 반복되는 질문과 데이터 요청에 대한 FAQ 페이지를 만들어 제공하여 응답 시간을 단축하고 부서원들이 SQL 에 능숙해지도록 교육 세션을 진행했습니다.

Wavve DATA



소유자: 노윤정 ***
안은영 (Deactivated)이(가) 5월 31, 2022에 마지막 업데이트

Terms

- **uno** : 이용자 개인 식별 번호 (aka **userId**)
- **hash uno** : 암호화 된 이용자 개인 식별 번호 (aka **sha512 uno** or **sha512 userId**)
- **profile** : 회원 정보
- **productCode** : 상품 (aka **pk**)
- **유료가입자** : uno 단위 (프리미엄 4회선 상품 = 1명)
- **유료이용자** : 회선 단위 (프리미엄 4회선 상품 = 4명)
 - **유료이용자(다회선 환산)** : 매출을 7900원 1인 기준으로 환산했을 때 몇 명인가
- **DW** : Data Warehouse란 사용자의 의사 결정에 도움을 주기 위하여 기간시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변환해서 관리하는 데이터베이스를 말한다.
- **DM** : Data Mart는 데이터 웨어하우스(Data Warehouse, DW) 환경에서 정의된 접근계층으로, 데이터 웨어하우스에서 데이터를 꺼내 사용자에게 제공하는 역할을 한다. 데이터 마트는 데이터 웨어하우스의 부분이며, 대개 특정한 조직, 혹은 팀에서 사용하는 것을 목적으로 한다.



데이터를 이해하고 문서화하며 데이터 요청을 관리하는 경험을 통해 도메인 관련 지식을 쌓을 수 있었습니다. 다양한 부서에서 데이터 추출 요청 업무를 전달받고 작업하는데 있어 궁금한 부분은 테이블 담당자에게 질문하면서 효과적인 데이터 관련 커뮤니케이션 방법을 학습하였습니다. 이러한 노력으로 데이터마케팅팀의 데이터 관리와 협업의 효율성을 향상시키고 조직 내 데이터를 향한 관심을 한층 더 높일 수 있었다고 생각합니다.

CRM을 통한 성과 지표 설계 및 사용자 세분화 분석

데이터 사이언티스트로서 마케팅팀 내에서 데이터 기반 의사 결정을 강화하기 위해 데이터를 제공하고 데이터 이해도를 향상시키는 역할을 수행했습니다. Tableau 대시보드를 활용하여 가입자의 코호트 분석을 시각화하여 사용자 세분화를 진행하였으며, 이탈률을 신속히 모니터링하여 사용자 앱 활성화를 촉진하기 위한 푸시 메시지 파이프라인을 개발하였습니다. 또한, 각 고객 그룹의 푸시 메시지 발송 성공율을 추적하기 위한 파이프라인을 Pentaho Data Integration을 활용하여 구축하고 결과 데이터를 몽고디비에 저장하였습니다. 더욱 흥미로운 푸시 메시지를 제공하기 위해 프로그램의 인기와 영향력을 정량화할 수 있는 지표가 필요하다고 판단하였고, B2C 구독자가 구독한 날에 시청한 프로그램을 집계하여 "First View"라는 성과 지표를 개발하였습니다.



"First View" 지표를 개발하는 과정에서, Pentaho Data Integration (또는 Kettle) 프로그램을 활용하여 매일 아침 8시에 실행되는 ETL 프로세스를 통해 B2C 유료 상품에 가입한 가입자들의 정보를 추출하였습니다. 이 데이터에서는 가입자가 가입한 날에 시청한 VOD와 MOVIE 프로그램의 아이디를 집계하여, 각 프로그램별로 유료 가입자가 시청한 유료 콘텐츠를 파악하였습니다. 이를 통해 어떤 프로그램이 유료 가입자가 서비스에 가입하게 된 원인인지를 수치화하여 프로그램의 인기를 평가하였습니다. 또한, 이 데이터는 엑셀 파일 형태로 매일 아침 8시에 리더직급에게 공유되었으며, 전일의 "First View" 데이터뿐만 아니라 이전 주의 동일 요일 데이터도 함께 제공하여 주 단위로 트렌드를 확인할 수 있도록 제공되었습니다.

각 프로그램의 "First View" 지표는 사용자 참여의 주요 지표로 작용하여, 사용자 페르소나 분석 및 오리지널 프로그램 제작을 위한 중요한 통찰력을 제공하였습니다. 이 데이터는 추후 매체전략부에서 각 부서별 KPI를 설정하는 데 큰 기여를 하게 되었습니다.