

Data Sense White Paper

Data as an Asset

(Data Governance & Quality Management)

Data Profiling & Quality Analysis

Physical & Logical Relationship Diagram

Value Chain & System Mapping

PII & Security Compliance Analysis

Date: 2026년 1월

Author: Qliker qliker@kakao.com

Version: ver 2.0

목차

Data Sense 소개서	3
도입 배경 (Problem Statement)	3
현재 조직이 겪는 현실적 문제	3
DataSense가 해결하는 핵심 질문	4
DataSense 적용 후 변화	4
DataSense가 추구하는 가치 (Core Value)	5
주요 기능	5
DataSense의 특장점	6
결언	7
DataSense 솔루션 프로그램 상세 설명서	8
1. Configuration Setup (설정 및 환경 관리)	8
2. Data Analyzer (Profiling)	9
3. Data Quality Information (데이터 품질 가시화)	9
4. 23_Physical & Logical (데이터 관계 자산화)	10
5. Value Chain & System Analysis (프로세스-시스템-데이터 분석)	13
6. PII - Personally Identifiable Information (개인정보 보호)	16
7. Data Consistency Analysis (데이터 일관성 분석)	18
8. Character Analysis (데이터 결함 정밀 진단)	19
💡 종합적인 비즈니스 가치 (Executive Summary)	21

Data Sense 소개서

Enterprise Data Quality & Relationship Intelligence Platform

DataSense는 원천 데이터 프로파일링부터 비즈니스 가치 사슬(Value Chain)과 시스템까지 연결하여 데이터의 **생성-흐름-품질**을 통합 관리하는 지능형 솔루션입니다.

"데이터는 비즈니스의 언어입니다." (Data as a Business Language) 즉 데이터의 흐름에서 비즈니스의 가치를 찾는다.

도입 배경 (Problem Statement)

- **데이터 파편화(가시성 부재):** 시스템별로 흩어진 데이터가 비즈니스의 어느 단계에서 쓰이는지 파악 불능
- **보이지 않는 품질(신뢰성 저하):** 인코딩 오류(한글 깨짐), 결측치 등 기술적 결함이 비즈니스 의사결정 저해 및 차세대 AI 시스템 적용에 장애
- **영향도 분석의 한계(추적의 한계):** 특정 테이블 변경 시 하위 시스템에 미치는 영향 추적이 복잡하고 불투명하여 시스템 변경 시 발생하는 영향도를 수작업으로 분석하여 시간과 비용 낭비

현재 조직이 겪는 현실적 문제

데이터 관점

- 테이블 수 증가 → 전체 구조 파악 불가
- ERD 문서는 노후화 또는 미존재

업무 관점

- 신규 분석/보고마다 구조 재해석 필요
- 담당자 변경 시 지식 단절 및 신규 시스템 도입에 기존 지식과 연결에 어려움

리스크 관점

- 컬럼 변경 → 영향 범위 예측 불가
- 잘못된 조인으로 품질 이슈 발생

✖ 데이터는 있지만 신뢰 가능한 구조 정보는 없음

비용으로 환산한 문제점 현황 (Hidden Cost)

구분	연간 비용 영향
분석 준비 시간	분석 인력의 30~50% 소모
구조 파악 회의	반복 커뮤니케이션 비용
장애/재작업	품질 오류로 인한 재분석
인수인계	개인 의존적 지식 손실

보이지 않지만 지속적으로 발생하는 구조 인식 비용

DataSense가 해결하는 핵심 질문

- 이 데이터는 어디서 생성되고 어디로 흐르는가?
- 컬럼 간 관계는 논리적으로 타당한가?
- 어떤 데이터가 **핵심(Value-Critical)** 인가?
- 데이터 품질 문제는 어디에서 발생하는가?

→ 기존에는 사람이 답하던 질문을 시스템이 즉시 시각화

DataSense 적용 후 변화

구조 인식

- 데이터 값 기반 **Logical ERD** 자동 생성
- 선택한 테이블 기준 관련 구조 즉시 확장

운영 방식

- ERD는 과거의 문서가 아닌 **항상 최신 상태의 결과물**
- 컬럼명 기반의 ERD가 아닌 데이터 기반의 논리적 ERD

사용자 경험

- 비개발자도 데이터 구조 이해 가능

✓ 데이터 구조가 자산이 되는 순간

DataSense가 추구하는 가치 (Core Value)

- **Visibility:** 데이터의 생성부터 소멸까지 투명한 흐름 제공
- **Accuracy:** 기술적/논리적 결함을 실시간 탐지하여 고품질 데이터 자산화
- **Alignment:** IT 인프라와 비즈니스 가치 사슬(Value Chain)의 완벽한 일치

주요 기능

① 지능형 데이터 프로파일링

- **자동화된 품질 진단:** Null 비율, 데이터 형식(Format) 준수율, 유니크 값 비율 자동 산출
- **기술 결함 탐지:** 한글 깨짐(Broken Kor), 유니코드 오류 등 데이터 이관 시 발생 가능한 위험을 선제적으로 식별
- **데이터 신뢰도 정량화:** 파일별 품질 상태를 정량화하여 관리 우선순위 도출
- **비즈니스 가치:** 데이터 자산 규모를 숫자로 인지하고 관리 대상과 우선순위를 명확화

② Pure Downstream 계보 분석

- **물리적 영향도 추적:** 특정 파일을 선택하면 컬럼 이름으로 연결된 하위 시스템을 즉시 분석 (Physical Relationship Diagram)
- **지능형 영향도 추적:** 특정 파일을 선택하면 **데이터 값으로** 연결된 하위 시스템을 즉시 분석 (Logical Relationship Diagram)
- **순환 참조 해소:** 상호 순환 참조(Mutual Recursion)를 자동으로 필터링하여 명확한 데이터 흐름(Lineage) 제공
- **시각적 가시성:** 색상 체계를 통해 영향도 수준(Direct vs. Extended) 구분

차별점

- 물리 ERD
- 실제 사용 관계 기반 논리 ERD

③ 비즈니스 가치 사슬(Value Chain) 매핑

- Value Chain의 Primary(주활동)와 Support(지원활동)로 구분된 산업별 Value Chain 정의
- **비즈니스-시스템 연결:** 비즈니스 활동에 필요한 시스템과 원천 파일을 매핑하여 각 Value Chain Activity가 어떤 IT 시스템 및 데이터 파일에 의존하는지 분석
- **상세 아키텍처 다이어그램:** 비즈니스 중심의 시스템 구조도(Summary/Detailed)를 자동으로 생성하여 요약(Summary) 모드 분석에서 파일 단위의 상세(Detailed) 모드 분석으로 심화
- 산업(Industry) → Activity → System → File 구조화

→ "어떤 데이터가 중요한가"에 대한 명확한 답

•

DataSense의 특장점

① 분석의 완결성

- **차트의 근거 제시 :** 시각적 차트 하단에 원천 데이터를 배치하여 "차트의 근거"를 즉시 확인
- **Recursive-Free 로직:** 복잡하게 꼬인 데이터 관계를 논리적으로 해소하여 분석 오류 방지
- **실시간 리포팅:** 데이터 변경 시 즉각적으로 반영되는 동적 리포팅 환경 제공

DataSense 적용의 기대 효과 (Expected Benefits)

- 데이터 품질 관리 자동화
- 시스템 간 데이터 흐름 가시화
- 데이터 표준/거버넌스 기반 강화
- 분석·AI 프로젝트의 신뢰도 향상
- 의사결정 속도 향상: 데이터 영향도 및 품질 분석 시간 **75% 이상 단축**
- 데이터 신뢰도 제고: 선제적 결함 탐지를 통한 고품질 데이터 자산 확보

- IT-비즈니스 정렬:** IT 부서와 비즈니스 부서 간 동일한 데이터 맵(Map)을 통한 협업 강화

항목	Before	After
구조 파악 시간	수일	수분
신규 분석 착수	반복 조사	즉시 가능
변경 영향 분석	경험 의존	ERD 기반
데이터 신뢰도	개인 판단	구조 기반

→ **분석 생산성 2~3배 향상 기대**

조직 관점의 전략적 가치

- 데이터 거버넌스 기반 확보로 AI 확산에 대응
- 분석/개발/업무 간 공통 언어 제공
- 특정 인력 의존도 감소

결언

- DataSense는 단순한 툴이 아닙니다. **비즈니스의 가시성을 확보하는 가장 강력한 전략적 자산입니다.**
- DataSense는 데이터를 관리 대상이 아니라 의사결정을 가속하는 구조 자산으로 전환합니다.
- DataSense는 단순 분석 도구를 넘어 데이터를 이해하는 프레임워크입니다
 - ✓ 구조를 알고
 - ✓ 품질을 확인하고
 - ✓ 가치를 연결합니다

기준 방식	DataSense 적용
문서 중심	데이터 중심
사람 중심	자동화된 구조 인식
사후 대응	선제적 영향 분석

🏛️ DataSense 솔루션 프로그램 상세 설명서

데이터 전문 솔루션인 DataSense의 각 프로그램을 비즈니스 가치와 운영 효율성 관점에서 상세히 설명해 드립니다. 이 설명서는 IT 자산이 어떻게 비즈니스 성과로 연결되는지를 중점적으로 다룹니다.

이 도구들은 단순한 기술 스택을 넘어, 기업의 데이터를 ****비즈니스 자산(Asset)****으로 전환하고 운영 리스크를 관리하는 핵심적인 역할을 수행합니다.

1. 📄 Configuration Setup (설정 및 환경 관리)

- 프로그램명: 01_Configuration_Setup.py
- 비즈니스 관점의 의미: 전사 데이터 분석을 위한 '표준 가이드라인' 설정 단계입니다.
- 주요 효과:
 - 거버넌스 체계 구축: 분석 대상이 되는 소스 경로와 시스템 범위를 정의하여 데이터 관리의 일관성을 확보합니다.
 - 운영 비용 절감: 환경 설정을 GUI 기반으로 관리함으로써, 개발자의 개입 없이 현업 관리자가 분석 범위를 조정할 수 있어 유지보수 효율이 향상됩니다.

Configuration Setup

환경 설정 편집기입니다.

⚙️ 환경설정 편집

상단에서 경로/파일을 조정한 뒤 로드하세요. 편집 후 저장을 누르면 해당 경로/파일로 저장됩니다.

✍️ 기본 항목

2. Data Analyzer (Profiling)

- 프로그램명: 11_Data Analyzer (Profiling).py
- 비즈니스 관점의 의미: 전사 데이터의 건강 상태를 진단하는 '정밀 건강검진' 도구입니다.
- 주요 효과:
 - 데이터 원천의 신뢰성 확보: 데이터의 타입, 통계적 분포, 관계를 프로파일링하여 분석의 기초가 되는 로우 데이터(Raw Data)의 정합성을 보장합니다.
 - 분석 준비 시간(Time-to-Insight) 단축: 수작업으로 진행하던 데이터 탐색 과정을 자동화하여, 분석가가 비즈니스 전략 수립에 집중할 수 있는 시간을 벌어줍니다.

Data Analyzer (Data Profiling)

데이터 프로파일링을 수행하여 데이터 분석을 위한 다양한 통계 정보를 생성합니다.

- 데이터 프로파일링을 수행하여 데이터 분석을 위한 다양한 통계 정보를 생성하고 (Data Quality Analyzer)
- 데이터 타입 및 사전 정의된 Rule을 적용하여 데이터의 의미를 분석하며 (Data Type & Rule Analyzer)
- 데이터 간의 논리적 관계도 정보를 생성합니다. (Data Relationship Analyzer)

통계 내역에 포함된 정보들은 다음과 같습니다.

속성 정보	Value 정보	Length 정보	Value 구성	Value Top 10	문자 통계
데이터 타입	Primary Key 여부	Length 종류	영문, 한글, 숫자 등 패턴 구성	Top 10 값	영문 대소문자 열 수
오리클 타입	데이터 값의 열 개수	Length 최소	패턴의 종류 수	Top 10 비율	한글 포함 열 수
를 기반 타입	Uniqueness 비율	Length 최대	다빈도 패턴 구성		숫자 포함 열 수
	Null 비율	Length 다빈도	다빈도 패턴 및 비율		특수문자 열 수
	최소/최대/평균/중앙값	Length 평균/중앙값			혼합 문자 열 수

▼  실행 비밀번호 입력

비밀번호를 입력하세요
?


분석 대상 파일의 수 및 크기에 따라 시간이 많이 소요될 수 있습니다. (약 10분 이상 소요)

 Data Analyzer 실행

3. Data Quality Information (데이터 품질 가시화)

- 프로그램명: 21_Data Quality Information.py
- 비즈니스 관점의 의미: 데이터 기반 의사결정의 **'신뢰도 보증 수표'**입니다.
- 주요 효과:

- 의사결정 리스크 감소: Null 비율, 값의 분포 등 통계 정보를 제공하여 부정확한 데이터로 인한 잘못된 경영 판단을 사전에 방지합니다.
- 데이터 정제 우선순위 도출: 품질이 낮은 데이터를 즉각 식별하여 리소스 투입의 효율성을 극대화합니다.

Data Quality Information

Data Analyzer의 결과를 기반으로 각 컬럼들에 대한 통계 정보입니다.

File Statistics

67 Data File #	162 만건 Total Record #	91 Bytes Total File Size	2026-01-19 Work Date
--------------------------	---------------------------------	------------------------------------	--------------------------------

Data Quality Information

아래의 탭에서 삼세 정보를 확인할 수 있습니다.

Value Info Value Type Info Top10 Info Length Info Character Info DQ Score Info Total Statistics

모든 컬럼들의 데이터 값 정보를 제공합니다.

FileName	ColumnName	OracleType	PK	ValueCnt	Null(%)	UniqueCnt	Unique(%)	MinString	MaxString	ModeString	MedianString	ModeCnt
COMPANY.csv	CompanyID	VARCHAR2(4)	1	1474	0	1474	100	0	9999	0	410	1
COMPANY.csv	CompanyType	VARCHAR2(1)	0	1474	0	9	0.61	0	9	1	1	1393
COMPANY.csv	Company_Name	VARCHAR2(19)	0	1474	0	1400	94.98	(유)컨버스코리아	충진플러스	0	(주)한독에프엔씨	19
COMPANY.csv	Company_Name_Abbreviation	VARCHAR2(16)	0	1474	0	1371	93.01	(주)골프존커머스	충화	0	아티스	24
COMPANY.csv	Business_Registration_Number	NUMBER(10)	0	1422	3.53	1235	86.85	0	8898701966	2068650913	2041947026	21
COMPANY.csv	Business_Type	VARCHAR2(19)	0	1440	2.31	360	25	0	화장품 도매업	제조, 도소매	제조	151
COMPANY.csv	Business_Item	VARCHAR2(33)	0	1437	2.51	814	56.65	0	화장품/집화	의류	의류	233

탭에서 다양한 정보들을 분석할 수 있습니다.

4. ④ 23_Physical & Logical (데이터 관계 자산화)

- 프로그램명: 22_Data Relationship Diagram.py & 23_Physical & Logical Diagram.py
- 비즈니스 관점의 의미: 기업의 복잡한 데이터 구조를 *****비즈니스 언어*****로 번역하는 지도입니다.
- 주요 효과:
 - 지식 자산의 표준화: 담당자 퇴사나 변경 시에도 비즈니스 로직(Logical)과 실제 데이터(Physical)의 관계를 즉시 파악할 수 있어 업무 인수인계 비용을 절감합니다.
 - 영향도 분석 가속화: 시스템 변경 시 어떤 데이터가 영향을 받는지 실시간으로 확인하여 장애 발생 확률을 대폭 낮춥니다.

데이터 관계 (Data Relationship Diagram)

Data Value Mapping 기반 논리적 Data Relationship Diagram을 생성합니다.

File Statistics

67

162 만건

91 Bytes

2026-01-19

Data File #

Total Record #

Total File Size

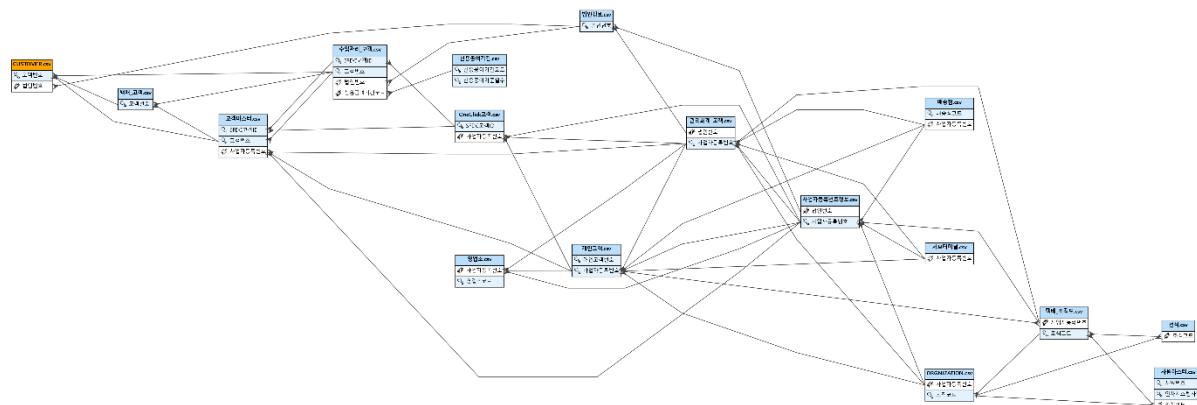
Work Date

1. 테이블 선택

선택	FileName	PK Columns	Column #	Max_Level	Rel Table #	FilePath
<input type="checkbox"/>	COMPANY.csv	CompanyID	22	4	6	C:/projects/myproject/QDQM/DS_Input/Internal/COMP
<input type="checkbox"/>	CONTRACT.csv	CONTRACT_NO	32	3	12	C:/projects/myproject/QDQM/DS_Input/Internal/CON
<input checked="" type="checkbox"/>	CUSTOMER.csv	고객번호	31	4	14	C:/projects/myproject/QDQM/DS_Input/Internal/CUS
<input type="checkbox"/>	ORGNIZATION.csv	조직코드	22	4	12	C:/projects/myproject/QDQM/DS_Input/Internal/ORG
<input type="checkbox"/>	OneClick고객.csv	SFDC고객ID	11	3	5	C:/projects/myproject/QDQM/DS_Input/Internal/OneC
<input type="checkbox"/>	PRM계약.csv	PRM계약번호	18	4	6	C:/projects/myproject/QDQM/DS_Input/Internal/PRM
<input type="checkbox"/>	test.csv	ID	21	3	5	C:/projects/myproject/QDQM/DS_Input/Internal/test
<input type="checkbox"/>	간선사업자.csv	사업자번호	6	0	0	C:/projects/myproject/QDQM/DS_Input/Internal/간선
<input type="checkbox"/>	개인고객.csv	개인고객번호, 사업자등	23	3	9	C:/projects/myproject/QDQM/DS_Input/Internal/개인

분석할 파일
선택

선택할 파일에 대한 컬럼명에 의한 물리적 데이터 관계도를 생성합니다.



데이터 관계도는 연결 깊이를 사용자가 선택합니다.

연결관계 정보 설정

연결관계 모드 선택

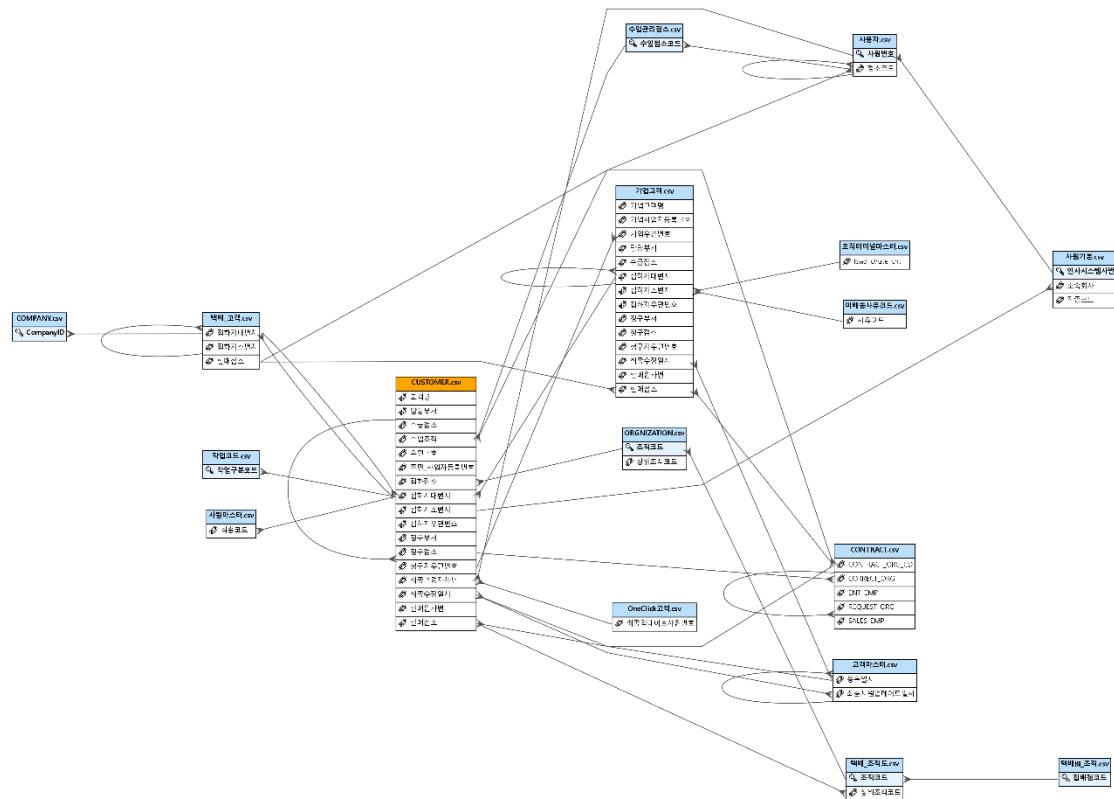
- Physical
- Logical
- Physical & Logical

탐색 깊이 (Level)

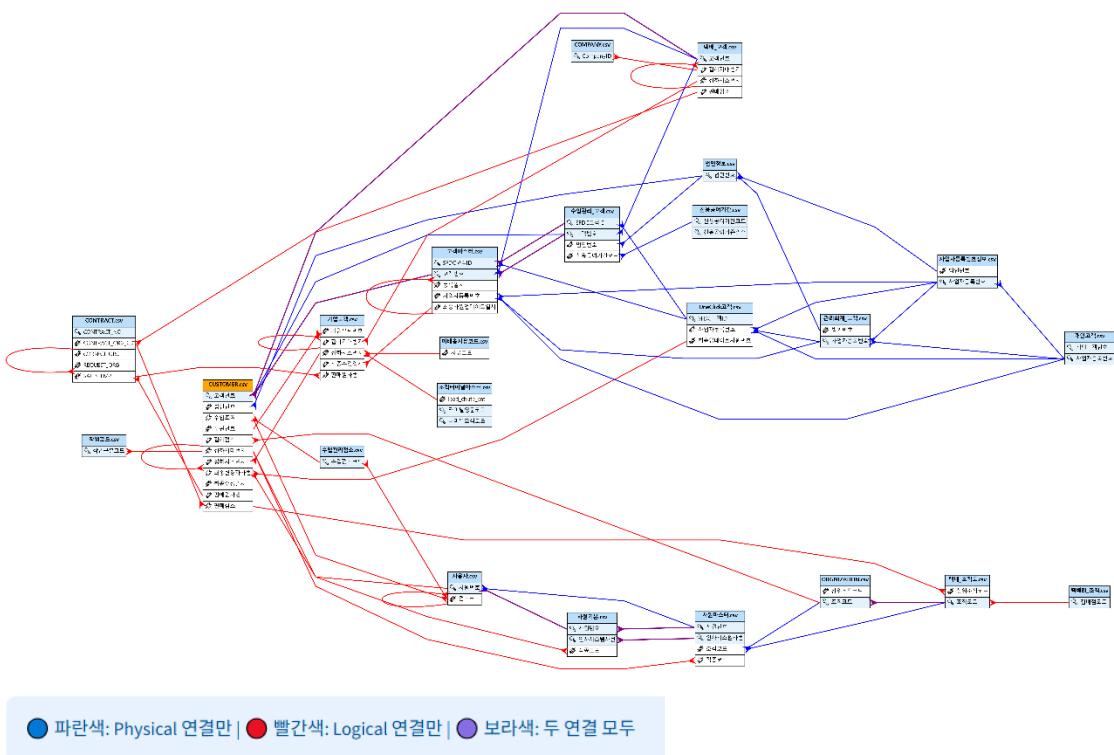
전체 컬럼 표시

4

선택할 파일에 대한 데이터 값에 의한 논리적 데이터 관계도를 생성합니다.



선택할 파일에 대한 물리적 데이터 관계도 및 논리적 데이터 관계도를 생성합니다.



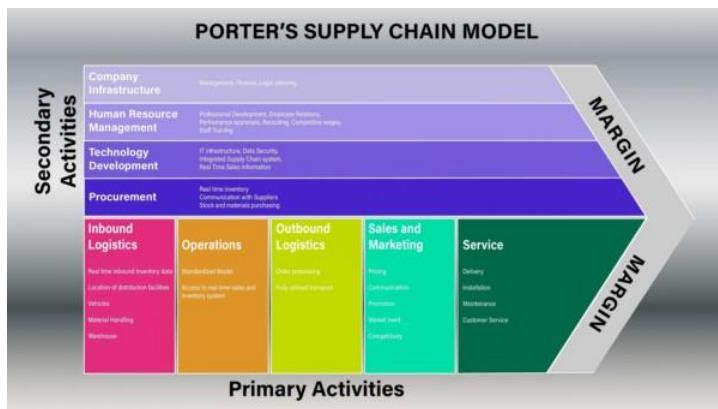
생성한 데이터 관계도를 구성하는 상세 정보를 분석합니다.

ERD 생성 정보 (컬럼 단위)

FileName	ColumnName	PK	FK	Physical_Type	Logical_Type	Both_Type	Check	Related_Table	Related_Column	Level	Displayed	Is_Anchor	Relationship_Type
COMPANY.csv	CompanyID	Y	N	None	None	None	None			0	Y	N	None
COMPANY.csv	CompanyID	Y	Y	0	1	0	1	택배_고객.csv	집까지대번지	0	Y	N	None
CONTRACT.csv	CONTRACT_NO	Y	N	0	0	0	0			0	Y	N	None
CONTRACT.csv	CONTRACT_ORG_CD	N	N	None	None	None	None			0	Y	N	None
CONTRACT.csv	CONTRACT_ORG_CD	N	Y	0	1	0	1	CUSTOMER.csv	판매점소	0	Y	N	None
CONTRACT.csv	CONTRACT_ORG_CD	N	Y	0	1	0	1	택배_고객.csv	판매점소	0	Y	N	None
CONTRACT.csv	CORRECT_ORG	N	N	None	None	None	None			0	Y	N	None
CONTRACT.csv	CORRECT_ORG	N	Y	0	1	0	1	CONTRACT.csv	REQUEST_ORG	0	Y	N	None
CONTRACT.csv	REQUEST_ORG	N	N	None	None	None	None			0	Y	N	None
CONTRACT.csv	REQUEST_ORG	N	Y	0	1	0	1	CONTRACT.csv	CORRECT_ORG	0	Y	N	None

5. Value Chain & System Analysis (프로세스-시스템-데이터 분석)

- 프로그램명: 26_Value Chain & System Diagram.py & 27_Value Chain & System Analysis.py
- 비즈니스 관점의 의미: IT 시스템과 '실제 매출이 발생하는 사업 영역' 사이의 연결 고리를 분석합니다.
- 주요 효과:
 - IT 자산 최적화: 특정 비즈니스 프로세스에 중복된 시스템이 있는지, 혹은 지원되지 않는 사각지대가 있는지 파악하여 IT 투자의 가시성을 높입니다.
 - 비즈니스 연속성(BCP) 확보: 특정 파일이나 시스템 장애 시 어떤 사업 부서가 타격을 입는지 즉시 파악하여 위기 대응 속도를 높입니다.



1 Select Industry

Registered Industry List

물류-택배

☞ 다음 탭을 순차적으로 수행하세요.

[Value Chain Definition](#) [System Definition](#) [Map Value Chain & System to File](#)

2 Value Chain Activity Definition

[Activity List \(Edit/Delete\)](#) [Add New Activity](#)

💡 설정 방법: 표 내부의 값을 직접 클릭하여 수정 후 하단 변경사항 저장 버튼을 클릭하세요.

순번은 중복이 되지 않도록 순차적으로 일련번호를 부여합니다.

번호	산업군	구분	순번	활동명(영문)	활동명(한글)	설명
11	물류-택배	Primary	1	Marketing & Sales	마케팅&영업	마케팅 활동을 정의합니다.
12	물류-택배	Primary	2	Inbound Logistics	집하	집하 활동을 정의합니다.
13	물류-택배	Primary	3	Operations	간선	간선 활동
14	물류-택배	Primary	4	Outbound Logistics	배송	배송 활동
15	물류-택배	Primary	5	Services	사후관리	서비스 관리
16	물류-택배	Support	6	Finance	재무	재무 관리
18	물류-택배	Support	7	HR	인적자원관리	인사/급여 관리
17	물류-택배	Support	8	Infrastructure	기업인프라	인프라 관리

Value Chain 의 주요 활동과 운영중인 시스템을 관리하여 각 파일을 매핑합니다. 이를 다이어그램으로 분석합니다.

3 Value Chain & System Diagram

Value Chain & System을 기반으로 Value Chain Diagram과 System Architecture Diagram을 생성합니다.

Industry Selection

분석할 산업군을 선택하세요

물류-택배

8

Activity #

6

System #

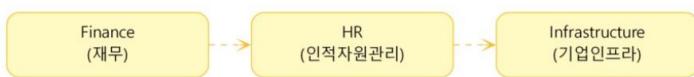
30

File #

조회할 탭을 선택하세요.

[Value Chain Diagram](#) [Value Chain & System Mapping](#) [Value Chain, System & File Mapping](#) [Value Chain Diagram & File Mapping](#)

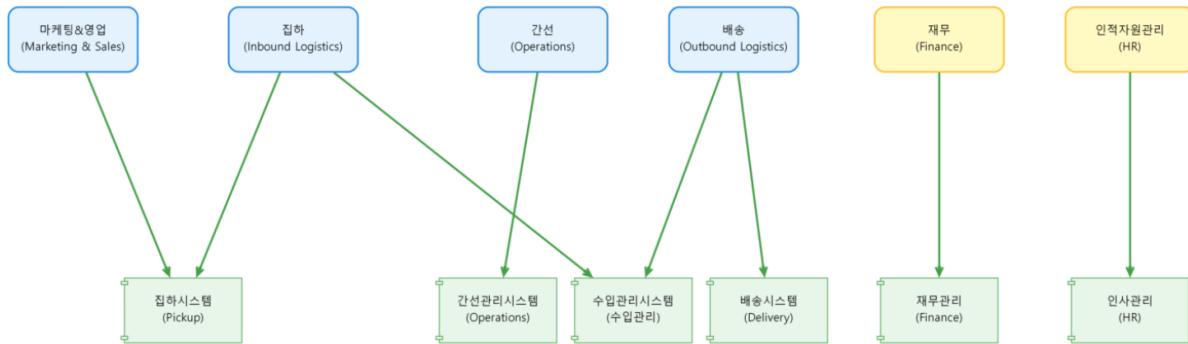
Value Chain Diagram (물류-택배)

> [Value Chain Data Preview](#)

 System Architecture (Summary): 물류-택배

Activity와 System 간의 연결 관계를 요약하여 표시합니다.

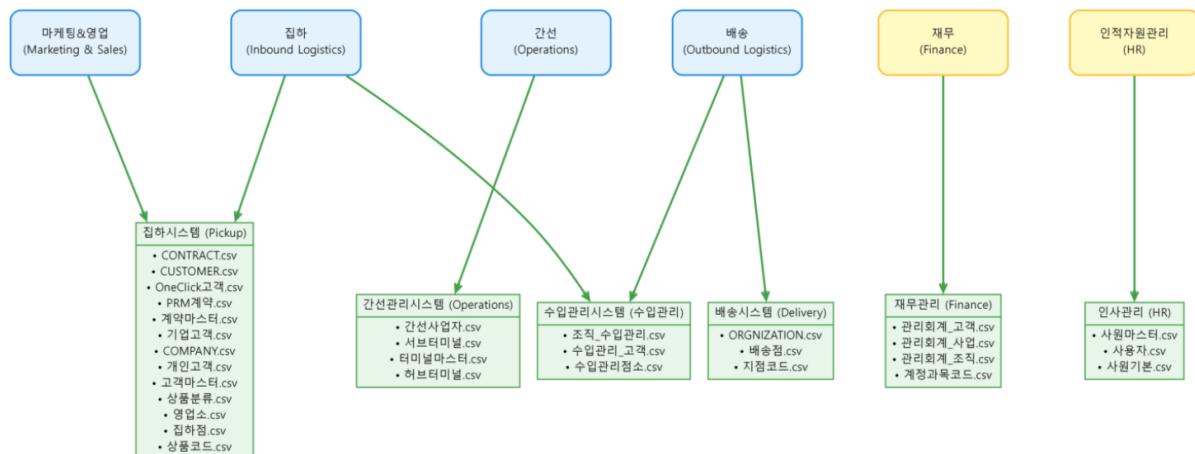
> System Architecture Data Preview



 System Architecture (Detail): 물류-택배

System 내부의 파일 목록을 포함한 상세 구성도를 표시합니다.

> System Architecture Data Preview



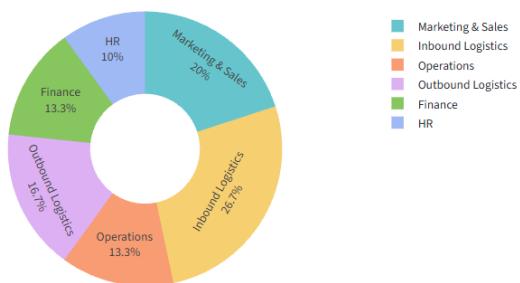
Activity Analysis

Activity별 파일 분포(파이 차트) Activity별 파일 수(막대 차트) Activity별 파일 수(테이블)

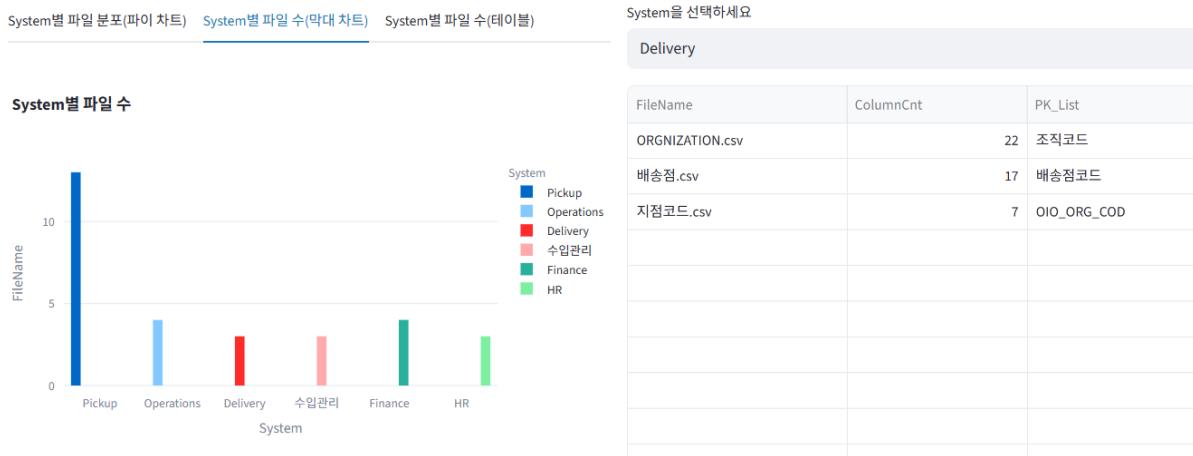
Activity를 선택하세요

Finance

Activity별 파일 분포



System Analysis



6. PII - Personally Identifiable Information (개인정보 보호)

- 프로그램명: 29_PII-Personally Identifiable Information.py
 - 비즈니스 관점의 의미: 기업의 **법적 리스크 방어막**이자 고객 신뢰 유지의 핵심 도구입니다.
 - **주요 기능:** 데이터 관계도(ERD) 위에 개인정보(이름, 주민번호, 연락처 등)의 분포와 참조 현황을 시각화합니다.
 - 주요 효과:
 - 컴플라이언스 자동화: 수만 개의 컬럼 중 개인정보를 자동으로 식별하고 관계를 추적하여 개인정보보호법 준수 비용을 획기적으로 낮춥니다.
 - 보안 사고 예방: 데이터 흐름상 보안이 취약한 지점을 시각화하여 데이터 유출 사고 발생 가능성을 선제적으로 차단합니다.
 - **비즈니스 효과:**
 - 컴플라이언스 대응 가속화: 개인정보 보호법(GDPR, 개인정보보호법 등) 준수 여부를 즉각적으로 점검하여 법적 리스크를 최소화합니다.
 - 데이터 보안 거버넌스 강화: 어떤 시스템의 어떤 테이블을 통해 개인정보가 흘러가는지 추적(Lineage)함으로써, 데이터 접근 제어 정책 수립 시 명확한 근거를 제공합니다.

PII(Personally Identifiable Information)

데이터 관계도 위에 개인정보(PII) 분포와 참조 현황을 시각화하여 보안 취약점을 식별합니다.

개인 식별 정보(PII, Personally Identifiable Information)란 이름, 주민등록번호, 주소, 전화번호, 이메일 등 개인을 직접적으로 또는 다른 정보와 결합하여 식별할 수 있는 모든 데이터를 말하며,

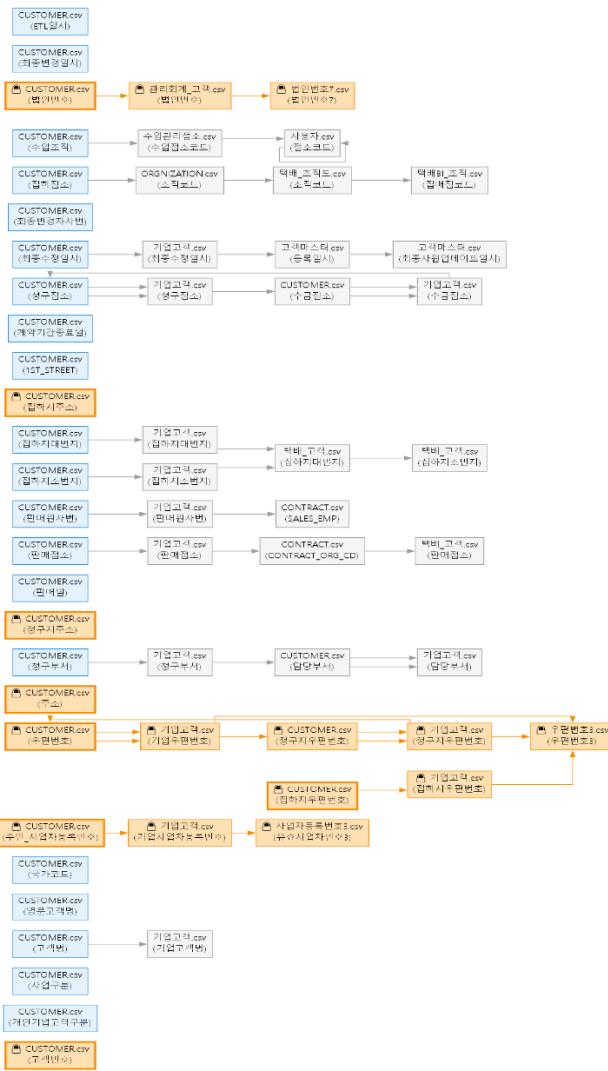
보안 진단 대상 설정

진단할 파일을 선택하세요

CUSTOMER.csv

🛡 보안 오버레이 실행

선택된 파일을 기준으로 PII 맵을 생성합니다.



주황색 노드는 개인정보 포함 가능성이 높은 컬럼입니다. 마스킹 및 접근 권한을 점검하십시오.

7. Data Consistency Analysis (데이터 일관성 분석)

- 프로그램명: 32_Data Consistency Analysis.py
- 비즈니스 관점의 의미: 여러 시스템에 흩어진 데이터의 '단일 진실 원천(Single Source of Truth)' 확보입니다.
- 주요 효과:
 - 부서 간 소통 오류 제거: 부서별로 데이터가 달라 발생하는 논쟁을 없애고 통합된 수치로 소통하게 합니다.
 - 데이터 통합(MDM) 효율화: 차세대 시스템 구축이나 ERP 통합 시 데이터 정합성 확인 시간을 단축합니다.

동일 컬럼이 시스템 전체에 Oracle Data Type 및 저장된 데이터의 값의 형태가 일치여부를 분석합니다. 이를 통하여 동일한 컬럼명이 다르게 사용되는 현황을 파악하고 데이터의 일치성을 분석합니다.

데이터 모델 일관성 분석

OracleType 불일치 Format 불일치

동일 컬럼명 내 OracleType이 다른 사례: 58건

ColumnName	OracleType	LenCnt	Count	FileList
고객번호	VARCHAR2(14)	1	1	급변회계_고객.csv
고객구분코드	VARCHAR2(8)	1	1	OneClick고객.csv
고객구분코드	VARCHAR2(8)	2	1	고객마스터.csv
고객번호	VARCHAR2(10)	1	2	고객마스터.csv, 수입관리_고객.csv
고객번호	VARCHAR2(7)	1	2	CUSTOMER.csv, 택배_고객.csv
국내집배구역	NUMBER(6)	2	1	사원마스터.csv
국내집배구역	VARCHAR2(7)	2	1	사원기본.csv
귀속부서조직코드	VARCHAR2(7)	3	3	관리회계QLIK_조직.csv, 관리회계_조직그룹상세.csv, 원가인명정보.csv
귀속부서조직코드	VARCHAR2(7)	4	1	관리회계BI_조직.csv
귀속부서조직코드	VARCHAR2(7)	5	1	관리회계_조직_test.csv
대사업명	VARCHAR2(7)	3	1	사업코드관리.csv

➊ 데이터 모델 일관성 분석

⚠️ OracleType 불일치 ⚠️ Format 불일치

FormatCnt 3이하 컬럼 중 Format 불일치: 58건

ColumnName	OracleType	Format	FormatCnt	FormatTop10	Count	FileList
고객구분코드	VARCHAR2(8)	AAAAAAA	2	"AAAAAAA", "AAAAAAA"	1	고객마스터.csv
고객번호	VARCHAR2(10)	nnnnnnnnnn	1	"nnnnnnnnnn"	2	고객마스터.csv, 수입관리_고객.csv
고객번호	VARCHAR2(7)	nnnnnnn	1	"nnnnnnn"	1	택배_고객.csv
고객번호	VARCHAR2(7)	nnnnnnn	2	"nnnnnnn", "Annnnnn"	1	CUSTOMER.csv
국내집배구역	NUMBER(6)	nnnnnn	2	"nnnnnn", "nnnnnn"	1	사원마스터.csv
국내집배구역	VARCHAR2(7)	nnnnnn	2	"nnnnnn", "nnnnnn"	1	사원기본.csv
귀속부서조직GROUP코드	VARCHAR2(10)	AAsAA	4	"AAsAA", "AAsAAA", "AAsAAAAAA", "AAsAsnnnn	1	관리회계_조직그룹상세.csv
귀속부서조직GROUP코드	VARCHAR2(10)	AAsAAAA	4	"AAsAAAA", "AAsAAAAA", "AAsA", "AAsAsnnnn	1	관리회계_조직그룹.csv
귀속부서조직코드	VARCHAR2(7)	AAAAAA	4	"AAAAAA", "AAAAA", "AAAAAA", "AAAnnn"	1	관리회계_조직그룹상세.csv
귀속부서조직코드	VARCHAR2(7)	AAAAA	5	"AAAAAA", "AAAAA", "AAAAAA", "AAAnnn", "AAAAAA	1	원가기여정보.csv
대사업명	VARCHAR2(7)	KKKKKK	3	"KKKKKK", "KKKKKK", "KKKK"	1	사업코드관리.csv
대사업명	VARCHAR2(7)	KKKKKK	4	"KKKKKK", "KKKKKK", "KKKK", "KK"	1	관리회계BI_사업.csv
등록사용자ID	VARCHAR2(7)	AAA	3	"AAA", "aaa", "nnnnnnn"	1	택배_조치도.csv

❷ 8. Character Analysis (데이터 결함 정밀 진단)

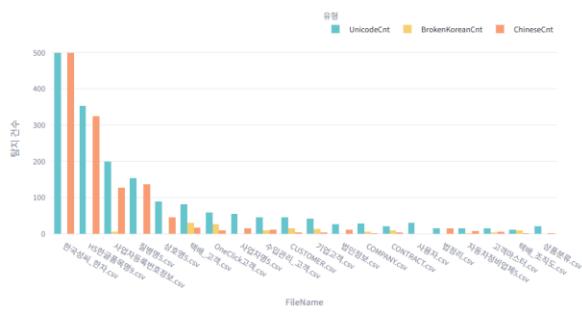
- 프로그램명: 34_Character Analysis.py
- 비즈니스 관점의 의미: 데이터 이관 및 연계 시 발생하는 ****'기술적 병목 현상'****을 사전에 해결합니다.
- 주요 효과:
 - 시스템 가동률 향상: 깨진 문자나 유니코드 및 특수문자로 인한 배치 작업 오류를 방지하여 시스템 운영 안정성을 확보합니다.
 - 글로벌 비즈니스 지원: 유니코드, 한자 등 다국어 데이터의 품질을 관리하여 글로벌 확장 시 데이터 호환성 문제를 해결합니다.

❸ Character & Quality Visualizer

비정형 문자(Unicode), 한자, 깨진 한글) 탐지 현황 및 상세 분석

분석 대상 파일	깨진 한글	한자(Chinese)	일본어(Japanese)
165개	154	1,279	58

▶ 파일별 결함 탐지 분포



	FileName	UnicodeCnt	BrokenKoreanCnt	ChineseCnt	TotalCnt
160	한국성씨_한자.csv	499	0	499	998
16	HS한글품목명9.csv	352	0	324	676
84	사업자등록번호정보.csv	200	5	127	332
145	질병명5.csv	153	0	136	289
92	상호명5.csv	90	0	45	135
149	택배_고객.csv	81	31	17	129
18	OneClick고객.csv	59	27	9	95
85	사업자명5.csv	55	0	16	71
100	수입관리_고객.csv	45	9	12	66
10	CUSTOMER.csv	45	15	4	64

파일을 선택하여 상세내역을 분석합니다.

파일별 상세 결합 내역

분석 결과를 확인할 파일을 선택하세요.

사업자등록번호정보.csv

컬럼별 통계 상세 레코드(Detail)

사업자등록번호정보.csv 의 컬럼별 탐지 현황

FileName	ColumnName	UnicodeCnt	BrokenKoreanCnt	ChineseCnt	JapaneseCnt	SpecialCnt	UnicodeChars	UnicodeOrdValues
사업자등록번호정보	사업자등록번호	0	0	0	2	None		None
사업자등록번호정보	고객명	200	5	127	58	75419	" ちばアイウエキグコシジスセタツテデトナニハバビホボミムヤユラ!	0x2018, 0x2019, 0x2161, 0x3000, 0x30
사업자등록번호정보	법인번호	0	0	0	0	6	None	None

상세레코드 탭을 선택하면 결합 데이터가 속한 행의 번호 및 상세 데이터값을 분석합니다. 이를 별도의 Cleansing 과정을 수행하여 정확한 데이터를 관리할 수 있습니다.

컬럼별 통계 상세 레코드(Detail)

사업자등록번호정보.csv 의 실제 결합 데이터 (최대 1,000건)

FileName	ColumnName	RecNo	CellValue	UnicodeChars	UnicodeOrdValues	UnicodeOrdInt	BrokenKoreanChars	ChineseChars	JapaneseChars
사업자등록번호정보.csv	고객명	59	중국民用航空局	中国局心民清用空算!	0x4e2d, 0x56fd, 0x5c40, 0x1 20013 22269 23616 245	None		中国局心民清用空算	None
사업자등록번호정보.csv	고객명	134	株式会社キヤスト キストヤ会式株社	0x30ad, 0x30b9, 0x30c8, 0x1 12461 12473 12488 125	None		会式株社	キストヤ	
사업자등록번호정보.csv	고객명	416	関西エアポート エアトボー会式株社	0x30a2, 0x30a8, 0x30a8, 0x1 12450 12456 12489 125	None		会式株社西關	エアトボー	
사업자등록번호정보.csv	고객명	934	수목공(秀木工)	工木秀	0x5de5, 0x6728, 0x79c0	24037 26408 31168	None	工木秀	None
사업자등록번호정보.csv	고객명	1492	上海外航服务公	上人公分力务司外服	0x4e0a, 0x4eba, 0x516c, 0x1 19978 20154 20844 205	None	上人公分力务司外服	None	
사업자등록번호정보.csv	고객명	1653	WING례제	㈜	0x321c	12828	None	None	None
사업자등록번호정보.csv	고객명	2291	가인(加人)	人加	0x4eba, 0x52a0	20154 21152	None	人加	None
사업자등록번호정보.csv	고객명	2755	河南省机场集团	公南司团场有机河省	0x516c, 0x5357, 0x53f8, 0x1 20844 21335 21496 222	None	公南司团场有机河省	None	
사업자등록번호정보.csv	고객명	4199	코센(季)	㈜	0x321c	12828	None	None	None
사업자등록번호정보.csv	고객명	4937	한국 텔레비전	한국 텔레비전	0xffffe, 0xffff, 0xffff7, 0xffff8	65390 65391 65399 654	None	한국 텔레비전	한국 텔레비전

💡 종합적인 비즈니스 가치 (Executive Summary)

DataSense 솔루션은 단순히 코드를 실행하는 도구가 아니라, **데이터를 신뢰 가능한 자산으로 전환**하는 플랫폼입니다.

1. 위험 관리(Risk Management): PII 분석을 통한 보안 사고 예방 및 법적 규제 준수.
2. 효율성(Efficiency): 비즈니스-시스템 매핑을 통한 IT 운영 최적화 및 커뮤니케이션 단축.
3. 수익성(Profitability): 고품질 데이터를 기반으로 한 정확한 분석과 전략 수립 지원.