|  |
| --- |
| Katta ma'lumotlar va Vizualizatsiya |
| Abdulloh Halilov |

**5-bo'lim: Katta ma'lumotlar va Vizualizatsiya**

**O‘qituvchi:** Behzod Qurbonov

**Guruh identifikatori:** 23-412

**Talaba ID:** 231397

**Taqdim etilgan sana:** 27.01.2025 yil

**BTEC o'quvchilar topshiriqlarini baholash va deklaratsiya**

Baholash uchun ishlarni taqdim etganda, har bir o'quvchi ish o'ziniki ekanligini tasdiqlovchi deklaratsiyani imzolashi kerak.

|  |  |
| --- | --- |
| **O‘quvchi (talaba) identifikatori:** | 231397 |
| **Baholovchi nomi:** | Behzod Qurbonov |
| **Birlik yoki komponent raqami va nomi:** | UNIT 5: Katta ma'lumotlar va Vizualizatsiya |
| **Topshiriq nomi:** | Internal Assignment |
| **Topshiriq topshirilgan sana:** | 27.01.2025yil |

Iltimos, har bir topshiriq uchun berilgan ishlarni sanab o'ting. Ishlarni topish mumkin bo'lgan sahifa raqamlarini ko'rsating yoki ishlarning mohiyatini tavsiflang (masalan, grafik, rasm).

|  |  |  |
| --- | --- | --- |
| **Topshiriq vazifasi ma'lumoti** | **Ishlar taqdim etildi** | **Sahifa** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
| **O'quvchi deklaratsiyasi**  Men ushbu topshiriq uchun taqdim etilgan ish meniki ekanligini tasdiqlayman. Ishda foydalanilgan manbalarga aniq havola qildim. Men noto'g'ri deklaratsiya noto'g'ri ishlashning bir shakli ekanligini tushunaman.  **O‘quvchi imzosi:**  **Sana:** 27.01.2025 yil |

Mundarija

[KIRISH 3](#_Toc188846244)

[1. Katta ma’lumotlarning asosiy tushunchalarini tushuntiring(tadqiq eting). 4](#_Toc188846245)

[2. Yakuniy foydalanuvchilar va tashkilotlarga qaratilgan qaror qabul qilish uchun ma'lumotlarning qiymatini baholang. 6](#_Toc188846246)

[3. Tashkilot uchun ma'lumotlarga asoslangan qaror qabul qilishning afzalliklari va muammolarini tahlil qiling 7](#_Toc188846247)

[4. Qaror qabul qilish uchun ma'lumotlardan foydalanishda ma'lumotlarning foydalanuvchilar va tashkilotlarga potentsial ta'sirini baholang. 9](#_Toc188846248)

[5. Sanoatda qo'llaniladigan katta ma'lumotlar va vizualizatsiya uchun statistik va grafik usullarni tavsiflang. 11](#_Toc188846249)

[6. Ma'lumotlarni tahlil qilish va vizualizatsiya qilish uchun mavjud bo'lgan turli sohadagi yetakchi vositalar va dasturiy yechimlarni ko'rib chiqing. 12](#_Toc188846250)

[7. Ma'lumotlarni tahlil qilish va vizualizatsiya qilish uchun turli sohadagi yetakchi vositalar va dasturiy yechimlardan qanday foydalanilishini misollar bilan solishtiring. 14](#_Toc188846251)

[8. O‘zingizning ma’lumotlarni tayyorlash va manipulyatsiya qilish mahoratingizni baholang, mavjud ma'lum ma'lumotlar to'plami uchun manfaatdor tomonlarning ehtiyojlarini qanday qondirishini ko‘rsatish uchun statistik usullarni tanlaganingizni asoslang. 16](#_Toc188846252)

[9. Laboratoriya ishi (Ma’lumotlar bazasi o‘qituvchi tomonidan taqdim etiladi): 23](#_Toc188846253)

[10. Ma’lumotlar bo'yicha mutaxassislar duch keladigan turli rollar, mas'uliyat va muammolarni tushuntiring. 41](#_Toc188846254)

[11. Ma'lumotlar muvofiqligini ta'minlash uchun ma'lumotlar mutaxassislari tomonidan qo'llaniladigan turli strategiyalarni ko'rib chiqing 42](#_Toc188846255)

[12. Ma’lumotlarga asoslangan madaniyatga axloqni shakllantirishda ma'lumotlar mutaxassislari duch keladigan rol, mas'uliyat va muammolarni tahlil qiling. 45](#_Toc188846256)

[Xulosa 47](#_Toc188846257)

[ADABIYOTLAR RO'YXATI 48](#_Toc188846258)

KIRISH

Zamonaviy biznes muhitida ma’lumotlar muhim ahamiyatga ega bo‘lib, tashkilotlarning muvaffaqiyatini belgilovchi asosiy omillardan biriga aylandi. Xususan, avtomobil sotuvi kabi raqobatbardosh sohada ma’lumotlarga asoslangan qarorlar qabul qilish kompaniyalarga operatsion samaradorlikni oshirish, mijozlar bilan munosabatlarni yaxshilash va raqobatda ustunlikka erishish imkonini beradi. Katta ma’lumotlarning paydo bo‘lishi va ularni tahlil qilishning yangi usullarining rivojlanishi bu borada yangi imkoniyatlar yaratdi. Ushbu ish avtomobil sotuvi bilan shug‘ullanuvchi kompaniyada katta ma’lumotlarni tahlil qilishning nazariy va amaliy jihatlarini ko‘rib chiqishni maqsad qiladi. Ish davomida quyidagi masalalar tahlil qilinadi: katta ma’lumotlarning asosiy tushunchalari va xususiyatlari; katta ma’lumotlar bilan ishlash bosqichlari va tashkilotda ularni boshqarish strategiyalari; yakuniy foydalanuvchilar va tashkilotlar uchun qaror qabul qilishda ma’lumotlarning qiymatini baholash; tashkilot uchun ma’lumotlarga asoslangan qaror qabul qilishning afzalliklari va muammolari; qaror qabul qilish uchun ma’lumotlardan foydalanishda ma’lumotlarning foydalanuvchilar va tashkilotlarga potensial ta’siri; sanoatda qo‘llaniladigan katta ma’lumotlar va vizualizatsiya uchun statistik va grafik usullar; ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun mavjud bo‘lgan turli sohadagi yetakchi vositalar va dasturiy yechimlar; ma’lumotlarni tayyorlash va manipulyatsiya qilish mahorati, manfaatdor tomonlarning ehtiyojlarini qondirish uchun statistik usullarni tanlash; ma’lumotlarga asoslangan madaniyatga axloqni shakllantirishda ma’lumotlar mutaxassislarining roli, mas’uliyati va muammolari. Ushbu ishning maqsadi avtomobil sotuvi bilan shug‘ullanuvchi kompaniyalarga katta ma’lumotlardan samarali foydalanish va operatsion samaradorlikni oshirish bo‘yicha amaliy tavsiyalar berishdan iborat. Ishda nazariy manbalar, ilmiy tadqiqotlar va amaliy misollar tahlilidan foydalaniladi. Kelgusi bo‘limlarda yuqorida keltirilgan masalalar batafsil ko‘rib chiqiladi.

1. Katta ma’lumotlarning asosiy tushunchalarini tushuntiring(tadqiq eting).

Avtomobil sotuvi bilan shug'ullanuvchi kompaniyada operatsion samaradorlikni oshirish uchun katta ma'lumotlarni tahlil qilishi lozimdur.

Ushbu ish doirasida avtomobil sotuvi bilan shug'ullanuvchi kompaniyadagi ma'lumotlar tahlilchisi rolida katta ma'lumotlarning nazariy asoslari, tahlil usullari va vizualizatsiyasi operatsion samaradorlikni oshirish uchun ko'rib chiqiladi.

1. Katta ma'lumotlarning fundamental tushunchalari

Katta ma'lumotlar (Big Data) – bu shunchalik katta hajmdagi, tez generatsiya qilinadigan va xilma-xil ma'lumotlar massivlariki, an'anaviy qayta ishlash usullari samarasiz bo'lib qoladi. Ular bir qator muhim xususiyatlar bilan tavsiflanadi, ular orasida eng mashhuri Dag Leyni tomonidan taklif qilingan 5V konsepsiyasidir:

Hajm (Volume): Terabaytlar, petabaytlar va hatto eksabaytlar bilan o'lchanadigan ulkan ma'lumotlar hajmi. Avtomobil kompaniyasi kontekstida bu sotuvlar, xizmat ko'rsatish, mijozlar, marketing kampaniyalari, avtomobillar datchiklaridan olingan ma'lumotlar va boshqalar haqidagi ma'lumotlar bo'lishi mumkin.

Tezlik (Velocity): Ma'lumotlarni generatsiya qilish va qayta ishlashning yuqori tezligi. Masalan, veb-saytga tashriflar, qo'ng'iroqlar markaziga qo'ng'iroqlar, real vaqt rejimida kelib tushadigan tranzaksiyalar haqidagi ma'lumotlar.

Xilma-xillik (Variety): Ma'lumotlarning turli formatlari: strukturalangan (ma'lumotlar bazalari), yarim strukturalangan (JSON, XML) va strukturalanmagan (matnlar, tasvirlar, videolar). Avtomobil biznesida bu mijozlarning matnli sharhlari, avtomobillarning fotosuratlari, videoshahlar va boshqalar bo'lishi mumkin.

Ishonchlilik (Veracity): Ma'lumotlarning noaniqligi va nomuvofiqligi, tozalash va to'g'rilikni tekshirishni talab qiladi. Masalan, mijozlar tomonidan ma'lumotlarni kiritishdagi xatolar, yozuvlarning dublikatlari.

Qiymat (Value): Ma'lumotlardan foydali bilimlar va tushunchalarni olish qobiliyati, ular qaror qabul qilish va biznesni yaxshilash uchun ishlatilishi mumkin.

2. Katta ma'lumotlar bilan ishlash bosqichlari

Katta ma'lumotlar bilan ishlash quyidagi bosqichlarni o'z ichiga oladi:

Ma'lumotlarni yig'ish (Data Collection): Turli manbalardan ma'lumotlarni olish: CRM-tizimlari, veb-saytlar, ijtimoiy tarmoqlar, avtomobillar datchiklari, yetkazib beruvchilarning ma'lumotlar bazalari va boshqalar.

Ma'lumotlarni filtrlash (Data Filtering): Tegishli ma'lumotlarni tanlash va keraksiz ma'lumotlarni o'chirish.

Ma'lumotlarni qayta ishlash (Data Processing): Ma'lumotlarni tahlil qilish uchun qulay formatga aylantirish, shu jumladan tozalash, normallashtirish va agregatsiya.

Ma'lumotlarni tozalash (Data Cleaning): Xatolarni tuzatish, dublikatlarni va zid ma'lumotlarni o'chirish.

Ma'lumotlarni vizualizatsiya qilish (Data Visualization): Tahlil qilish va tushunishni osonlashtirish uchun ma'lumotlarni grafik shaklda taqdim etish.

Ma'lumotlarni tahlil qilish (Data Analysis): Qonuniyatlar va tendentsiyalarni aniqlash uchun statistik usullar, mashinaviy o'qitish va boshqa analitik vositalarni qo'llash.

So'rovlar (Queries): Aniq ma'lumotlarni olish uchun ma'lumotlarga so'rovlar shakllantirish.

Ma'lumotlardan foydalanish (Data Usage): Olingan bilimlarni qaror qabul qilish, biznes-jarayonlarni optimallashtirish va natijalarni yaxshilash uchun qo'llash.

Ma'lumotlarni saqlash (Data Storage): Hajm, kirish tezligi va xavfsizlik talablarini hisobga olgan holda ma'lumotlarni saqlashni tashkil etish.

3. Tashkilotda katta ma'lumotlarni boshqarish strategiyalari

Katta ma'lumotlarni samarali boshqarish uchun tashkilotlar turli strategiyalardan foydalanadilar, jumladan:

Ma'lumotlar ko'li (Data Lake): Boshlang'ich formatdagi ma'lumotlarning markazlashtirilgan ombori, strukturalangan va strukturalanmagan ma'lumotlarni saqlash imkonini beradi. Afzalliklari: moslashuvchanlik, katta hajmdagi ma'lumotlarni saqlash imkoniyati, turli xil ma'lumot turlarini qo'llab-quvvatlash. Kamchiliklari: boshqarish qiyinligi, malakali mutaxassislarga ehtiyoj.

Ma'lumotlar ombori (Data Warehouse): Tahliliy so'rovlar uchun mo'ljallangan strukturalangan ma'lumotlar ombori. Ma'lumotlar oldindan tozalanadi va yagona formatga aylantiriladi. Afzalliklari: tahliliy so'rovlarni bajarishda yuqori unumdorlik, foydalanish qulayligi. Kamchiliklari: cheklangan moslashuvchanlik, strukturalanmagan ma'lumotlarni saqlash qiyinligi.

Avtomobil kompaniyasi kontekstida ma'lumotlar ko'li barcha kelayotgan ma'lumotlarni, shu jumladan avtomobillar datchiklaridan olingan ma'lumotlar, mijozlar sharhlari, qo'ng'iroqlar yozuvlari va boshqalarni saqlash uchun ishlatilishi mumkin. Ma'lumotlar ombori ma'lumotlar ko'li asosida yaratilishi va sotuvlar, xizmat ko'rsatish, mijozlar va boshqa asosiy ko'rsatkichlar haqidagi strukturalangan ma'lumotlarni o'z ichiga olishi mumkin.

2. Yakuniy foydalanuvchilar va tashkilotlarga qaratilgan qaror qabul qilish uchun ma'lumotlarning qiymatini baholang.

Ma'lumotlarning qiymati ularning yakuniy foydalanuvchilar va tashkilotlar uchun qaror qabul qilish jarayoniga qo'shadigan hissasi bilan belgilanadi. Avtomobil sotuvi bilan shug'ullanuvchi kompaniya kontekstida ma'lumotlarning qiymatini quyidagi jihatlar orqali baholash mumkin:

Kompaniya tomonidan belgilangan istiqbolli rejalar asosida ishlash va qo'yilgan maqsadga erishish uchun mavjud ma'lumotlarning qiymatini baholash:

Mijozlar haqidagi ma'lumotlar: Mijozlarning demografik ma'lumotlari, xarid qilish tarixi, afzalliklari va boshqa ma'lumotlar kompaniyaga marketing kampaniyalarini maqsadli yo'naltirish, sotuvlarni oshirish va mijozlar bilan munosabatlarni mustahkamlashga yordam beradi. Masalan, ma'lumotlar asosida qaysi mijozlarning yangi avtomobilga qiziqishi borligini aniqlash va ularga shaxsiy takliflar yuborish mumkin.

Sotuvlar haqidagi ma'lumotlar: Sotuvlarning dinamikasi, eng ko'p sotiladigan modellar, sotuvlarning hududiy taqsimoti va boshqa ma'lumotlar kompaniyaga sotuv strategiyasini optimallashtirish, zaxiralarni boshqarish va kelgusi sotuvlarni prognoz qilishga yordam beradi.

Xizmat ko'rsatish haqidagi ma'lumotlar: Xizmat ko'rsatish markaziga murojaatlar soni, eng ko'p uchraydigan muammolar, mijozlarning qanoatlanish darajasi va boshqa ma'lumotlar kompaniyaga xizmat ko'rsatish sifatini oshirish, muammolarni tezkor hal qilish va mijozlarning loyalligini oshirishga yordam beradi.

Marketing kampaniyalari haqidagi ma'lumotlar: Marketing kampaniyalarining samaradorligi, reklama kanallarining ta'siri, mijozlarning reaktsiyasi va boshqa ma'lumotlar kompaniyaga marketing byudjetini optimallashtirish va eng samarali kanallarni aniqlashga yordam beradi.

Avtomobillarning texnik holati haqidagi ma'lumotlar: Avtomobillarning ishlashi, eskirishi, ta'mirlash tarixi va boshqa ma'lumotlar kompaniyaga texnik xizmat ko'rsatish jarayonlarini rejalashtirish va avtomobillarning ishonchliligini oshirishga yordam beradi.

Mavjud ma'lumotlarni tahlil qilish orqali qanday statistik natijalarga ega bo'lish mumkinligini baholash:

Deskriptiv statistika: Sotuvlar hajmi, o'rtacha narx, eng ko'p sotilgan model kabi ko'rsatkichlarni aniqlash.

Inferentsial statistika: Sotuvlar tendentsiyalarini aniqlash, kelgusi sotuvlarni prognoz qilish, marketing kampaniyalarining samaradorligini baholash.

Korrelatsiya va regressiya tahlili: Turli omillar (narx, reklama, mavsumiylik) va sotuvlar o'rtasidagi bog'liqlikni aniqlash.

Segmentatsiya tahlili: Mijozlarni demografik, xarid qilish tarixi va boshqa mezonlar bo'yicha segmentlarga ajratish.

3. Tashkilot uchun ma'lumotlarga asoslangan qaror qabul qilishning afzalliklari va muammolarini tahlil qiling

Ma'lumotlarga asoslangan qaror qabul qilish (Data-Driven Decision Making, DDDM) tashkilotlarga faktlar va dalillarga asoslangan holda qarorlar qabul qilish imkonini beradi. Bu usul taxminlar va intuitsiyaga tayangan holda qaror qabul qilishdan ko'ra samaraliroq hisoblanadi.

Ma'lumotlarga asoslangan qarorlar qabul qilishning afzalliklari:

Doimiy takomillashtirish va rejalashtirish (Continuous Improvement and Planning): Ma'lumotlar kompaniyaga o'z faoliyatini doimiy ravishda tahlil qilish, muammolarni aniqlash va ularni hal qilish yo'llarini topish imkonini beradi. Bu esa, o'z navbatida, doimiy takomillashtirishga olib keladi. Ma'lumotlar asosida kelgusi davrlar uchun aniq rejaлар tuzish, resurslarni samarali taqsimlash va maqsadларга erishish muddatlarini belgilash mumkin.

Hamkorlikdagi qarorlar (Collaborative Decisions): Ma'lumotлар turli bo'limlar va xodimlar o'rtasidagi hamkorlikni kuchaytiradi. Umumiy ma'lumotлар bazasi mavjudligi qarorlarni muhokama qilish va qabul qilish jarayonini shaffof va samarali qiladi.

Xarajatlarni kamaytirish (Cost Reduction): Ma'lumotlar samarasiz jarayonlarni aniqlash va ularni optimallashtirish orqali xarajatlarni kamaytirishga yordam beradi. Masalan, sotuvlar haqidagi ma'lumotlar asosida eng kam talab bo'lgan modellarni aniqlash va ularning ishlab chiqarilishini qisqartirish mumkin.

Real vaqtdagi tushunchalar (Real-Time Insights): Real vaqt rejimida kelib tushadigan ma'lumotlar kompaniyaga bozordagi o'zgarishlarga tezkor reaktsiya qilish va tegishli choralarni ko'rish imkonini beradi. Masalan, mijozlarning veb-saytga tashriflari haqidagi ma'lumotlar asosida marketing kampaniyalarini real vaqtda sozlash mumkin.

Yangi imkoniyatlarni baholash (Evaluating New Opportunities): Ma'lumotlar yangi bozorlar, mahsulotlar va xizmatlarni baholash uchun asos yaratadi. Masalan, bozor tadqiqotlari ma'lumotlari asosida yangi avtomobil modelini ishlab chiqarishning istiqbollarini baholash mumkin.

Raqamli savodxonlik (Data Literacy): Ma'lumotlarga asoslangan qarorlar qabul qilish xodimlarning raqamli savodxonligini oshirishni talab qiladi. Bu esa kompaniyaning umumiy intellektual salohiyatini oshirishga xizmat qiladi.

Ma'lumotlarga asoslangan madaniyatlar (Data-Driven Cultures): DDDM kompaniyada ma'lumotlarga asoslangan madaniyatni shakllantiradi. Bu madaniyatda qarorlar faktlar va dalillarga asoslanadi, taxminlar va intuitsiyaga emas.

Qarorlar qabul qilishdagi muammolar:

Nomuvofiq va nostandart ma'lumotlar (Inconsistent and Non-Standard Data): Turli manbalardan kelib tushadigan ma'lumotlar turli formatlarda bo'lishi va bir-biriga mos kelmasligi mumkin. Bu esa ma'lumotlarni tahlil qilish va ulardan to'g'ri xulosalar chiqarishni qiyinlashtiradi.

Tavsiflovchi va bashorat qiluvchi tendentsiyalar va ehtimollar bilan moslashtirish (Reconciling Descriptive and Predictive Trends and Probabilities): Ma'lumotlar nafaqat o'tgan davrni tavsiflashi, balki kelgusi tendentsiyalarni bashorat qilishi ham mumkin. Biroq, bashoratlarning aniqligi ma'lumotlarning sifati va tahlil usullariga bog'liq. Shuning uchun, tavsiflovchi va bashorat qiluvchi ma'lumotlarni to'g'ri moslashtirish va ulardan oqilona foydalanish muhimdir.

Ushbu afzalliklar va muammolarni hisobga olgan holda, kompaniyalar DDDM strategiyasini joriy etish va amalga oshirishga jiddiy e'tibor qaratishlari lozim. Bu esa ularga raqobatbardoshlikni oshirish, operatsion samaradorlikni yaxshilash va muvaffaqiyatga erishishga yordam beradi.

4. Qaror qabul qilish uchun ma'lumotlardan foydalanishda ma'lumotlarning foydalanuvchilar va tashkilotlarga potentsial ta'sirini baholang.

Ma’lumotlardan qaror qabul qilishda foydalanish foydalanuvchilar (xodimlar) va tashkilotga katta ta’sir ko‘rsatishi mumkin. Bu ta’sir ijobiy ham, salbiy ham bo‘lishi mumkin, shuning uchun uni to‘g‘ri baholash va boshqarish muhimdir.

Foydalanuvchilarga (xodimlarga) ta’siri:

Ijobiy ta’sir:

Qaror qabul qilishda samaradorlikning oshishi: Xodimlar aniq va dolzarb ma’lumotlarga ega bo‘lganda, ular tezroq va asoslangan qarorlar qabul qila oladilar. Bu esa ish jarayonlarining tezlashishiga va samaradorlikning oshishiga olib keladi.

Ish sifatining oshishi: Ma’lumotlar xodimlarga o‘z ishlarining natijalarini tahlil qilish va xatolarni tuzatish imkonini beradi. Bu esa ish sifatining oshishiga va mijozlarning qanoatlanishiga olib keladi.

Kasbiy rivojlanish: Ma’lumotlar bilan ishlash ko‘nikmalarini egallash xodimlarning kasbiy rivojlanishiga yordam beradi va ularning bozordagi qiymatini oshiradi.

Ishdan qanoatlanishning oshishi: Asoslangan qarorlar qabul qilish imkoniyati xodimlarning ishdan qanoatlanishini oshiradi va ularning motivatsiyasini kuchaytiradi.

Salbiy ta’sir:

Ma’lumotlarning ortiqcha yuklamasi (Information Overload): Ko‘p miqdordagi ma’lumotlar xodimlarni chalg‘itishi va qaror qabul qilish jarayonini qiyinlashtirishi mumkin.

Maxfiylik muammolari: Shaxsiy ma’lumotlarning noto‘g‘ri ishlatilishi maxfiylik muammolarini keltirib chiqarishi mumkin.

Ish o‘rinlarining qisqarishi xavfi: Avtomatlashtirilgan tahlil va qaror qabul qilish tizimlarining joriy etilishi ba’zi ish o‘rinlarining qisqarishiga olib kelishi mumkin.

Ma’lumotlarga qaramlik: Faqat ma’lumotlarga tayanib qaror qabul qilish inson omilini va intuitsiyani e’tiborsiz qoldirishi mumkin.

Tashkilotga ta’siri:

Ijobiy ta’sir:

Raqobatbardoshlikning oshishi: Ma’lumotlarga asoslangan qarorlar qabul qilish tashkilotga bozordagi o‘zgarishlarga tezkor reaksiya qilish va raqobatda ustunlikka erishish imkonini beradi.

Operatsion samaradorlikning oshishi: Ma’lumotlar biznes jarayonlarini optimallashtirish, xarajatlarni kamaytirish va resurslardan samarali foydalanishga yordam beradi.

Mijozlar bilan munosabatlarning yaxshilanishi: Mijozlar haqidagi ma’lumotlar kompaniyaga ularning ehtiyojlarini yaxshiroq tushunish va ularga moslashtirilgan xizmatlar ko‘rsatish imkonini beradi.

Xavflarni boshqarishning yaxshilanishi: Ma’lumotlar xavflarni aniqlash va ularni kamaytirish choralarini ko‘rishga yordam beradi.

Salbiy ta’sir:

Ma’lumotlarni yig‘ish va qayta ishlash xarajatlari: Katta hajmdagi ma’lumotlarni yig‘ish, saqlash va qayta ishlash tashkilot uchun katta xarajatlarni talab qilishi mumkin.

Ma’lumotlar xavfsizligi muammolari: Ma’lumotlarning yo‘qolishi yoki ularga ruxsatsiz kirish tashkilot uchun jiddiy muammolarni keltirib chiqarishi mumkin.

Texnologik qaramlik: Ma’lumotlar tahlili uchun yangi texnologiyalarga qaramlik tashkilotni texnologik o‘zgarishlarga zaif qilib qo‘yishi mumkin.

Ushbu ta’sirlarni hisobga olgan holda, tashkilotlar ma’lumotlardan qaror qabul qilishda foydalanish strategiyasini ishlab chiqishlari va amalga oshirishlari lozim. Bu strategiya ma’lumotlarning afzalliklaridan maksimal darajada foydalanish va salbiy ta’sirlarni minimallashtirishni ta’minlashi kerak.

5. Sanoatda qo'llaniladigan katta ma'lumotlar va vizualizatsiya uchun statistik va grafik usullarni tavsiflang.

Katta ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun sanoatda turli xil statistik va grafik usullar qo‘llaniladi. Ushbu usullar ma’lumotlardan ma’noli xulosalar chiqarish, tendensiyalarni aniqlash va qaror qabul qilishni qo‘llab-quvvatlash uchun ishlatiladi.

Statistik usullar:

Deskriptiv statistika (Descriptive Statistics): Ma’lumotlarni qisqacha tavsiflash uchun ishlatiladigan usullar. Bularga o‘rtacha qiymat (mean), mediana (median), moda (mode), standart og‘ish (standard deviation), dispersiya (variance) kabi ko‘rsatkichlar kiradi. Avtomobil sotuvi kontekstida bu usullar sotuvlar hajmini, o‘rtacha narxni, eng ko‘p sotilgan modelni va boshqa ko‘rsatkichlarni aniqlash uchun ishlatilishi mumkin.

Inferensiyal statistika (Inferential Statistics): Namuna ma’lumotlari asosida umumiy populyatsiya haqida xulosalar chiqarish uchun ishlatiladigan usullar. Bularga gipotezalarni tekshirish (hypothesis testing), regressiya tahlili (regression analysis), korrelyatsiya tahlili (correlation analysis) kabi usullar kiradi. Avtomobil sotuvi kontekstida bu usullar sotuvlar tendensiyalarini aniqlash, kelgusi sotuvlarni prognoz qilish, marketing kampaniyalarining samaradorligini baholash uchun ishlatilishi mumkin.

Segmentatsiya tahlili (Segmentation Analysis): Mijozlarni umumiy xususiyatlariga ko‘ra guruhlarga ajratish uchun ishlatiladigan usul. Bu usul marketing kampaniyalarini maqsadli yo‘naltirish va mijozlar bilan munosabatlarni yaxshilash uchun ishlatilishi mumkin.

Vaqt qatorlari tahlili (Time Series Analysis): Vaqt bo‘yicha ketma-ketlikda olingan ma’lumotlarni tahlil qilish uchun ishlatiladigan usul. Bu usul sotuvlarning mavsuniy o‘zgarishlarini aniqlash va kelgusi davrlar uchun prognozlar tuzish uchun ishlatilishi mumkin.

Mashinnoye obucheniye (Machine Learning): Ma’lumotlar asosida o‘rganadigan va bashoratlar qiladigan algoritmlar. Bu usul mijozlarning xatti-harakatlarini bashorat qilish, sotuvlarni prognoz qilish va boshqa muammolarni hal qilish uchun ishlatilishi mumkin.

Grafik usullar:

Gistrogramma (Histogram): Miqdoriy ma’lumotlarning taqsimotini ko‘rsatish uchun ishlatiladigan ustunli diagramma.

Chiziqli grafik (Line Chart): Vaqt bo‘yicha o‘zgarib turuvchi ma’lumotlarni ko‘rsatish uchun ishlatiladigan grafik.

Doiraviy diagramma (Pie Chart): Umumiyning qismlarini ko‘rsatish uchun ishlatiladigan diagramma.

Sochilish diagrammasi (Scatter Plot): Ikki o‘zgaruvchi o‘rtasidagi bog‘liqlikni ko‘rsatish uchun ishlatiladigan diagramma.

Issiqlik xaritasi (Heatmap): Ma’lumotlarning intensivligini ranglar yordamida ko‘rsatish uchun ishlatiladigan grafik.

Geografik xaritalar (Geographic Maps): Geografik ma’lumotlarni ko‘rsatish uchun ishlatiladigan xaritalar.

Ushbu statistik va grafik usullarni to‘g‘ri qo‘llash kompaniyaga katta ma’lumotlardan maksimal foyda olish va asoslangan qarorlar qabul qilish imkonini beradi. Masalan, sotuvlar haqidagi ma’lumotlarni tahlil qilish uchun vaqt qatorlari tahlili va chiziqli grafiklardan, mijozlarning segmentatsiyasi uchun esa klaster tahlili va doiraviy diagrammalardan foydalanish mumkin.

6. Ma'lumotlarni tahlil qilish va vizualizatsiya qilish uchun mavjud bo'lgan turli sohadagi yetakchi vositalar va dasturiy yechimlarni ko'rib chiqing.

Katta ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun turli sohalarda keng qo‘llaniladigan ko‘plab vositalar va dasturiy yechimlar mavjud. Ularni quyidagicha guruhlash mumkin:

1. Dasturlash tillari va kutubxonalar:

Python: Ma’lumotlarni tahlil qilish, mashinaviy o‘qitish va vizualizatsiya uchun eng mashhur dasturlash tillaridan biri. Pandas, NumPy, SciPy, Scikit-learn, Matplotlib, Seaborn, Plotly kabi kutubxonalar ma’lumotlarni qayta ishlash, statistik tahlil o‘tkazish va interaktiv grafik tasvirlar yaratish imkonini beradi.

R: Statistik hisoblash va grafik tasvirlar uchun maxsus yaratilgan dasturlash tili. ggplot2, dplyr, tidyr kabi paketlar ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun keng imkoniyatlar taqdim etadi.

SQL: Ma’lumotlar bazalari bilan ishlash uchun standart til. Ma’lumotlarni saralash, filtrlash, agregatsiya qilish va boshqa amallarni bajarish uchun ishlatiladi.

2. Vizualizatsiya vositalari:

Tableau: Interaktiv vizualizatsiya va biznes-analitika uchun kuchli vosita. Ma’lumotlarni turli xil grafik shakllarda tasvirlash, dashboardlar yaratish va ma’lumotlar bilan interaktiv ishlash imkonini beradi.

Power BI: Microsoft tomonidan ishlab chiqilgan biznes-analitika va interaktiv vizualizatsiya vositasi. Tableau kabi keng imkoniyatlarga ega va Microsoft ekotizimi bilan integratsiyalashgan.

Google Data Studio (Looker Studio): Bepul onlayn vizualizatsiya vositasi. Google servislari (Google Analytics, Google Sheets va boshqalar) bilan oson integratsiyalanadi.

Qlik Sense: Ma’lumotlar bilan assotsiativ ishlashga asoslangan biznes-analitika platformasi. Foydalanuvchilar ma’lumotlar o‘rtasidagi bog‘liqliklarni osongina aniqlashlari mumkin.

3. Katta ma’lumotlarni qayta ishlash platformalari:

Apache Hadoop: Katta hajmdagi ma’lumotlarni taqsimlangan holda qayta ishlash uchun mo‘ljallangan platforma. HDFS (Hadoop Distributed File System) ma’lumotlarni taqsimlangan holda saqlash imkonini beradi, MapReduce esa ma’lumotlarni parallel qayta ishlash uchun freymvorkdir.

Apache Spark: Katta ma’lumotlarni tezkor qayta ishlash uchun mo‘ljallangan platforma. Hadoop ga nisbatan tezroq ishlaydi va mashinaviy o‘qitish algoritmlarini qo‘llab-quvvatlaydi.

Apache Kafka: Real vaqt rejimida ma’lumotlar oqimini qayta ishlash uchun mo‘ljallangan platforma. Sensorlardan, ijtimoiy tarmoqlardan va boshqa manbalardan kelib tushadigan ma’lumotlarni qayta ishlash uchun ishlatiladi.

4. Bulutli platformalar:

Amazon Web Services (AWS): Amazon EMR (Hadoop va Spark uchun), Amazon Redshift (ma’lumotlar ombori uchun), Amazon QuickSight (vizualizatsiya uchun) kabi katta ma’lumotlarni qayta ishlash va tahlil qilish uchun turli xil xizmatlarni taqdim etadi.

Google Cloud Platform (GCP): Google Dataproc (Hadoop va Spark uchun), BigQuery (ma’lumotlar ombori uchun), Looker (vizualizatsiya uchun) kabi katta ma’lumotlarni qayta ishlash va tahlil qilish uchun turli xil xizmatlarni taqdim etadi.

Microsoft Azure: Azure HDInsight (Hadoop va Spark uchun), Azure Synapse Analytics (ma’lumotlar ombori uchun), Power BI (vizualizatsiya uchun) kabi katta ma’lumotlarni qayta ishlash va tahlil qilish uchun turli xil xizmatlarni taqdim etadi.

Avtomobil sotuvi bilan shug‘ullanuvchi kompaniya uchun eng mos vositalar va dasturiy yechimlar kompaniyaning ehtiyojlari, ma’lumotlar hajmi, budjet va mavjud IT infratuzilmasiga bog‘liq bo‘ladi. Masalan, kichik kompaniya uchun Power BI yoki Google Data Studio yetarli bo‘lishi mumkin, katta kompaniya uchun esa Hadoop, Spark yoki bulutli platformalardan foydalanish maqsadga muvofiq bo‘ladi. Python va R kabi dasturlash tillari esa ma’lumotlarni chuqur tahlil qilish va maxsus yechimlar yaratish uchun foydali bo‘ladi.

7. Ma'lumotlarni tahlil qilish va vizualizatsiya qilish uchun turli sohadagi yetakchi vositalar va dasturiy yechimlardan qanday foydalanilishini misollar bilan solishtiring.

Ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun turli sohalarda qo‘llaniladigan vositalar va dasturiy yechimlarning qo‘llanilishini misollar bilan solishtiraylik:

1. Sog‘liqni saqlash:

Muammo: Kasalliklarning tarqalish tendensiyalarini aniqlash, bemorlarning holatini kuzatish va davolash usullarining samaradorligini baholash.

Qo‘llaniladigan vositalar:

R: Statistik tahlil va bioinformatika uchun keng qo‘llaniladi. Epidemiologik tadqiqotlar, genetik ma’lumotlarni tahlil qilish va klinik sinovlar natijalarini qayta ishlash uchun ishlatiladi.

Python (Pandas, SciPy): Katta hajmdagi tibbiy ma’lumotlarni qayta ishlash va tahlil qilish uchun ishlatiladi.

Tableau/Power BI: Kasalliklarning tarqalish xaritalarini yaratish, bemorlarning holati dinamikasini ko‘rsatish va davolash usullarining samaradorligini vizualizatsiya qilish uchun ishlatiladi.

Misol: R yordamida diabet kasalligining tarqalish tendensiyalarini tahlil qilish va xavf omillarini aniqlash. Tableau yordamida esa bu ma’lumotlarni geografik xaritada vizualizatsiya qilish va xavf guruhlarini ko‘rsatish mumkin.

2. Moliya:

Muammo: Bozor tendensiyalarini prognoz qilish, xavflarni baholash, firibgarlikni aniqlash va mijozlarning xatti-harakatlarini tahlil qilish.

Qo‘llaniladigan vositalar:

Python (Pandas, NumPy, Scikit-learn): Moliyaviy ma’lumotlarni qayta ishlash, vaqt qatorlarini tahlil qilish, risklarni baholash va mashinaviy o‘qitish algoritmlarini qo‘llash uchun ishlatiladi.

SQL: Tranzaksiyalar ma’lumotlar bazalari bilan ishlash uchun ishlatiladi.

Tableau/Power BI: Moliyaviy ko‘rsatkichlarni vizualizatsiya qilish, bozor tendensiyalarini ko‘rsatish va xavflarni baholash uchun ishlatiladi.

Misol: Python yordamida aksiyalar narxlarining vaqt qatorlarini tahlil qilish va kelgusi narxlarni prognoz qilish. Tableau yordamida esa bu ma’lumotlarni grafik shaklda vizualizatsiya qilish va treyderlarga qaror qabul qilishda yordam berish mumkin.

3. Riteyl (chakana savdo):

Muammo: Mijozlarning xatti-harakatlarini tahlil qilish, sotuvlarni prognoz qilish, marketing kampaniyalarining samaradorligini baholash va tovarlarni boshqarish.

Qo‘llaniladigan vositalar:

SQL: Sotuvlar, mijozlar va tovarlar haqidagi ma’lumotlar bazalari bilan ishlash uchun ishlatiladi.

Python (Pandas, Scikit-learn): Mijozlarni segmentatsiya qilish, sotuvlarni prognoz qilish va marketing kampaniyalarining samaradorligini baholash uchun ishlatiladi.

Tableau/Power BI: Sotuvlar dinamikasini, mijozlar segmentlarini va marketing kampaniyalarining natijalarini vizualizatsiya qilish uchun ishlatiladi.

Misol: Python yordamida mijozlarni xarid qilish tarixi va demografik ma’lumotlari asosida segmentlarga ajratish. Tableau yordamida esa bu segmentlarni grafik shaklda vizualizatsiya qilish va har bir segment uchun moslashtirilgan marketing strategiyalarini ishlab chiqish mumkin.

4. Avtomobil sotuvi:

Muammo: Sotuvlarni prognoz qilish, mijozlarning ehtiyojlarini aniqlash, marketing kampaniyalarining samaradorligini baholash va xizmat ko‘rsatish sifatini oshirish.

Qo‘llaniladigan vositalar:

SQL: Sotuvlar, mijozlar, avtomobillar va xizmat ko‘rsatish haqidagi ma’lumotlar bazalari bilan ishlash uchun ishlatiladi.

Python (Pandas, Scikit-learn): Sotuvlarni prognoz qilish, mijozlarni segmentatsiya qilish, marketing kampaniyalarining samaradorligini baholash va avtomobillarning texnik holatini tahlil qilish uchun ishlatiladi.

Tableau/Power BI: Sotuvlar dinamikasini, mijozlar segmentlarini, marketing kampaniyalarining natijalarini va xizmat ko‘rsatish ko‘rsatkichlarini vizualizatsiya qilish uchun ishlatiladi.

Misol: Python yordamida mijozlarning qiziqishlari va xarid qilish tarixi asosida ularga mos keladigan avtomobil modellarini tavsiya qilish. Tableau yordamida esa bu tavsiyalarni interaktiv katalog shaklida vizualizatsiya qilish va mijozlarga taqdim etish mumkin.

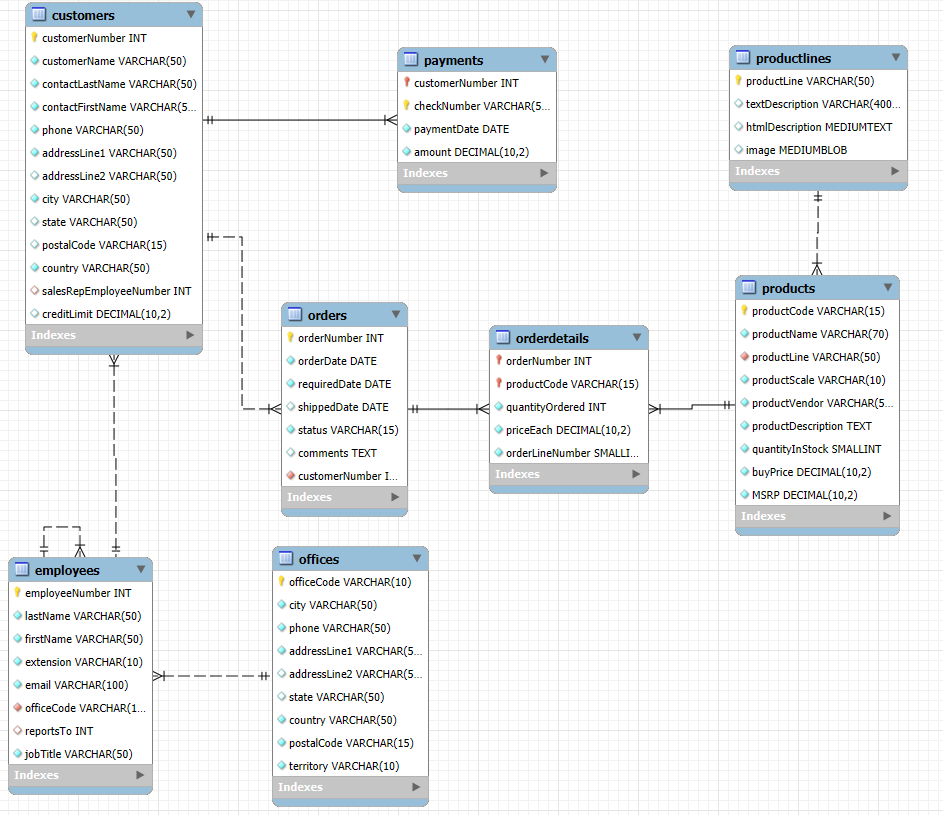
Ushbu misollar turli sohalarda ma’lumotlarni tahlil qilish va vizualizatsiya qilish uchun qanday vositalar va usullar qo‘llanilishini ko‘rsatadi. Har bir sohaning o‘ziga xos xususiyatlari va muammolari mavjud bo‘lib, shunga mos ravishda eng samarali vositalar va usullar tanlanadi.

8. O‘zingizning ma’lumotlarni tayyorlash va manipulyatsiya qilish mahoratingizni baholang, mavjud ma'lum ma'lumotlar to'plami uchun manfaatdor tomonlarning ehtiyojlarini qanday qondirishini ko‘rsatish uchun statistik usullarni tanlaganingizni asoslang.

Ma’lumotlarni tayyorlash va manipulyatsiya qilish – bu tahlil uchun xom ma’lumotlarni tozalash, o‘zgartirish va tashkil qilish jarayoni. Bu jarayon tahlilning muvaffaqiyati uchun juda muhim, chunki sifatli ma’lumotlargina ishonchli natijalar beradi.

Ma’lumotlarni tayyorlash va manipulyatsiya qilish bosqichlari:

Ma’lumotlarni yig‘ish (Data Collection): Turli manbalardan ma’lumotlarni to‘plash:



Ushbu Table refrencingga asoslanib biz tablelarni birlashtirishni boshlaymiz. Buning uchun birinchi navbatda barcha tablelarni birlashtirish uchun query yozamiz:

A screenshot of a computer program

AI-generated content may be incorrect.

Natija:

A screenshot of a computer

AI-generated content may be incorrect.

Ko’rinib turibdiki Columnlar juda ko’p, bularni anglash qiyin. Shuning uchun Data cleaning qismimizda alohida alohida tablelarni tuzatib, tozalab chiqamiz.

Ma’lumotlarni tozalash (Data Cleaning): Xatolar, dublikatlar va nomuvofiqliklarni aniqlash va tuzatish:

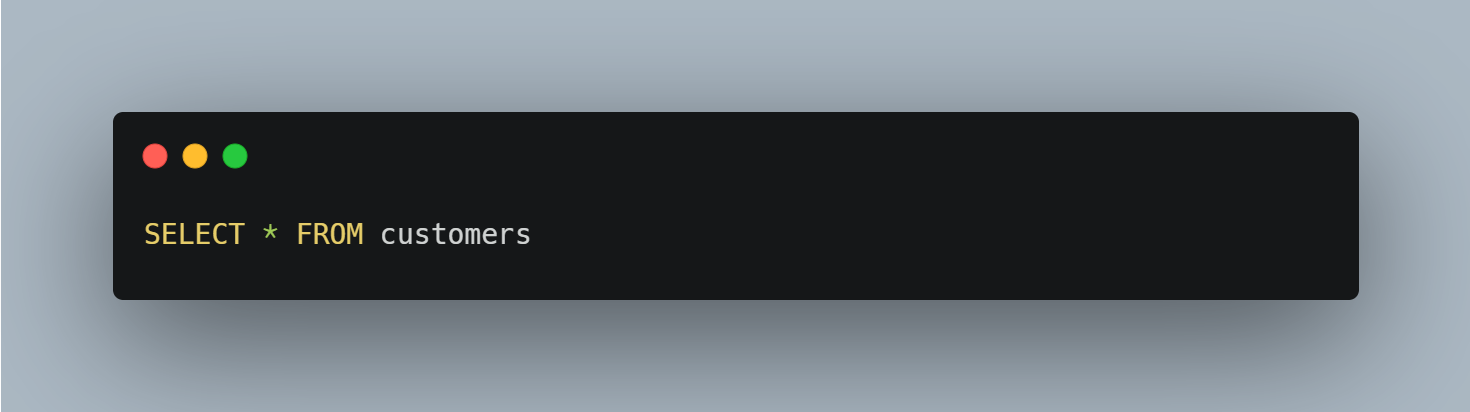
Ma’lumotlarni o‘zgartirish (Data Transformation): Ma’lumotlarni tahlil uchun qulay formatga keltirish.

Ma’lumotlarni integratsiya qilish (Data Integration): Turli manbalardan kelgan ma’lumotlarni birlashtirish.

Ma’lumotlarni qisqartirish (Data Reduction): Tahlil uchun keraksiz ma’lumotlarni olib tashlash.

Manfaatdor tomonlarning ehtiyojlarini qondirish uchun statistik usullarni tanlash:

Tozalashni birinchi Customerlardan boshlaymiz:

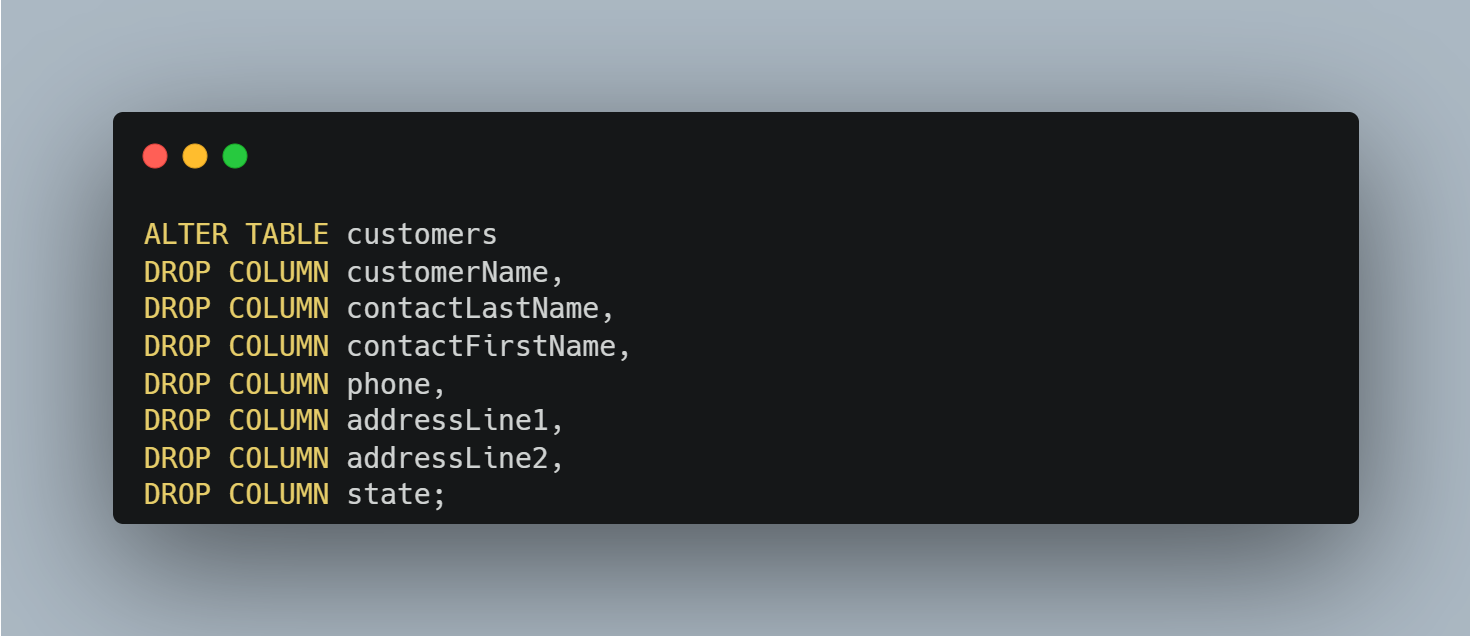


Natija:

A screenshot of a computer

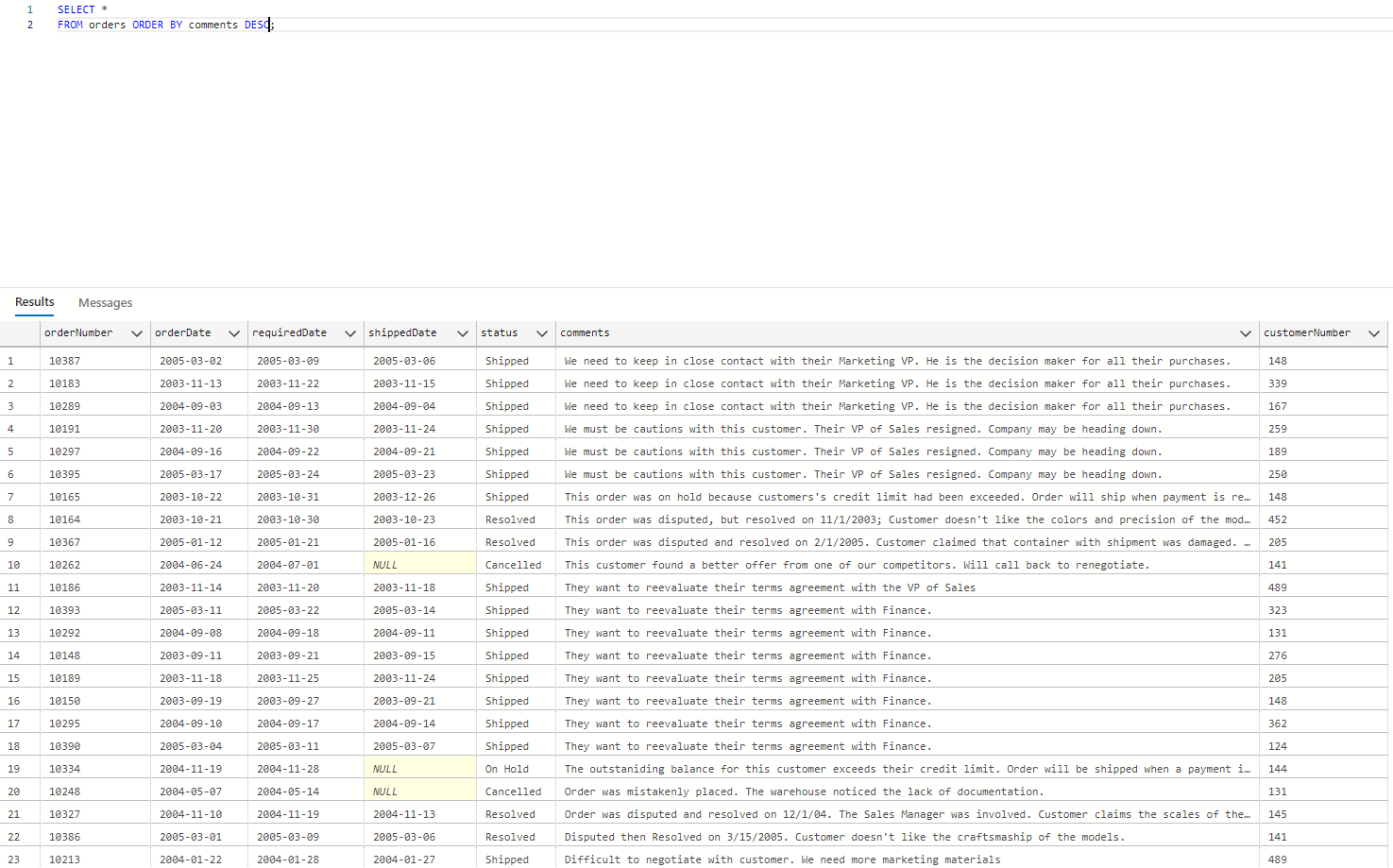
AI-generated content may be incorrect.

Bu yerda biz ishlatmaydigan ko’plab ma’lumotlar bor. Bularni tozalaymiz. Telefon raqamlari va shunga o’xshash columnlarni tuzatsa bo’ladi lekin biz manipulatsiya jarayonida bundan foydalanmaymiz. Shuning uchun ushbu columnni ham olib tashlaymiz:

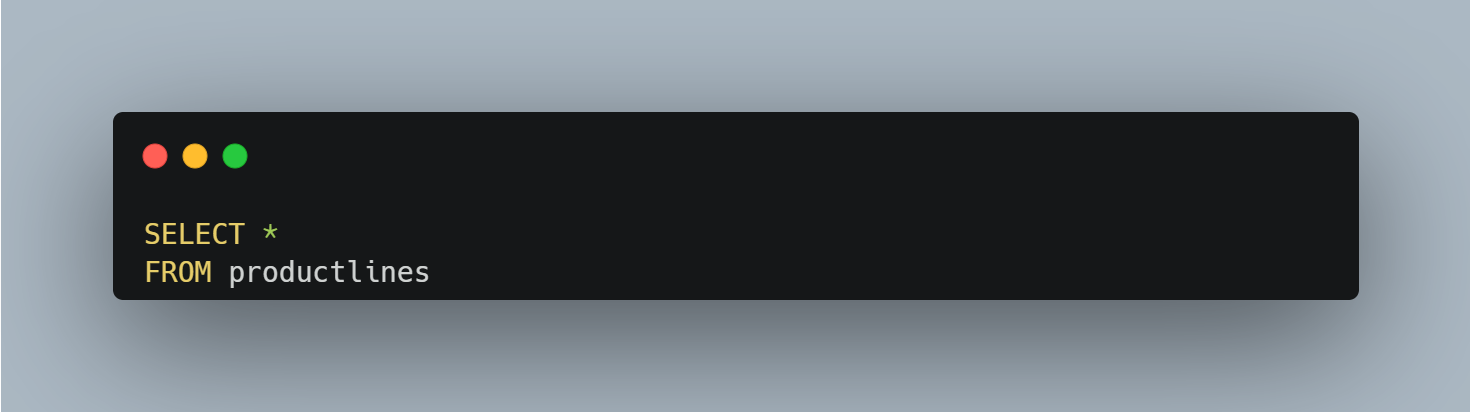


Ishlatilmaydigan tablelar olib tashlandi.

Orders Tablega o’zgartirish kiritilmadi, Comments bilan tanishib qichildi.



Productlines Table:

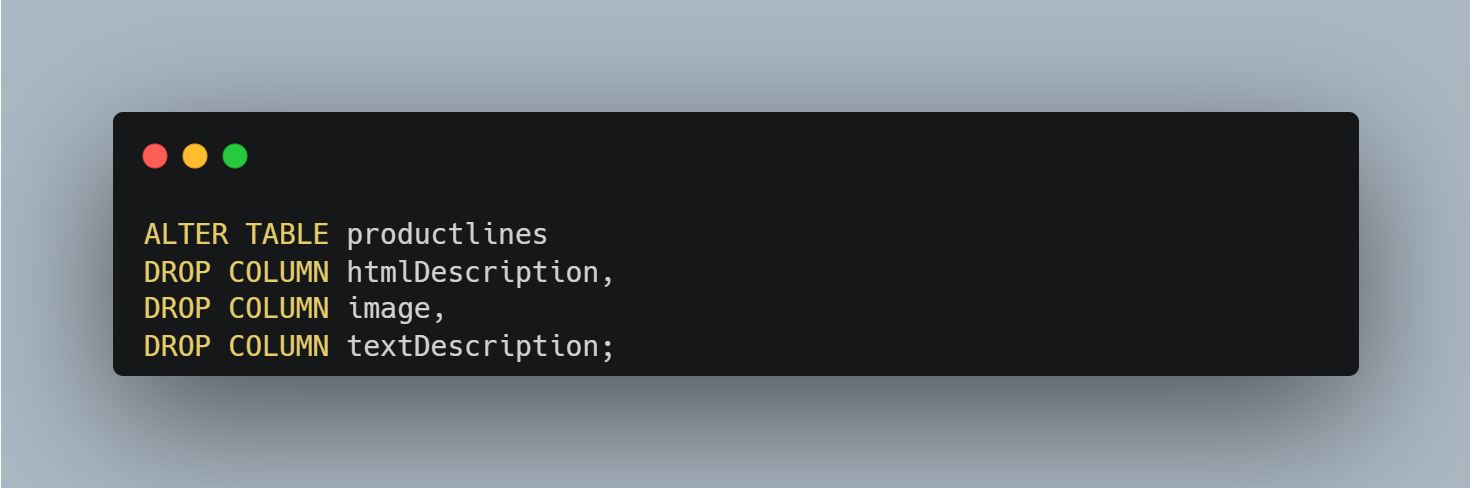


Natija:

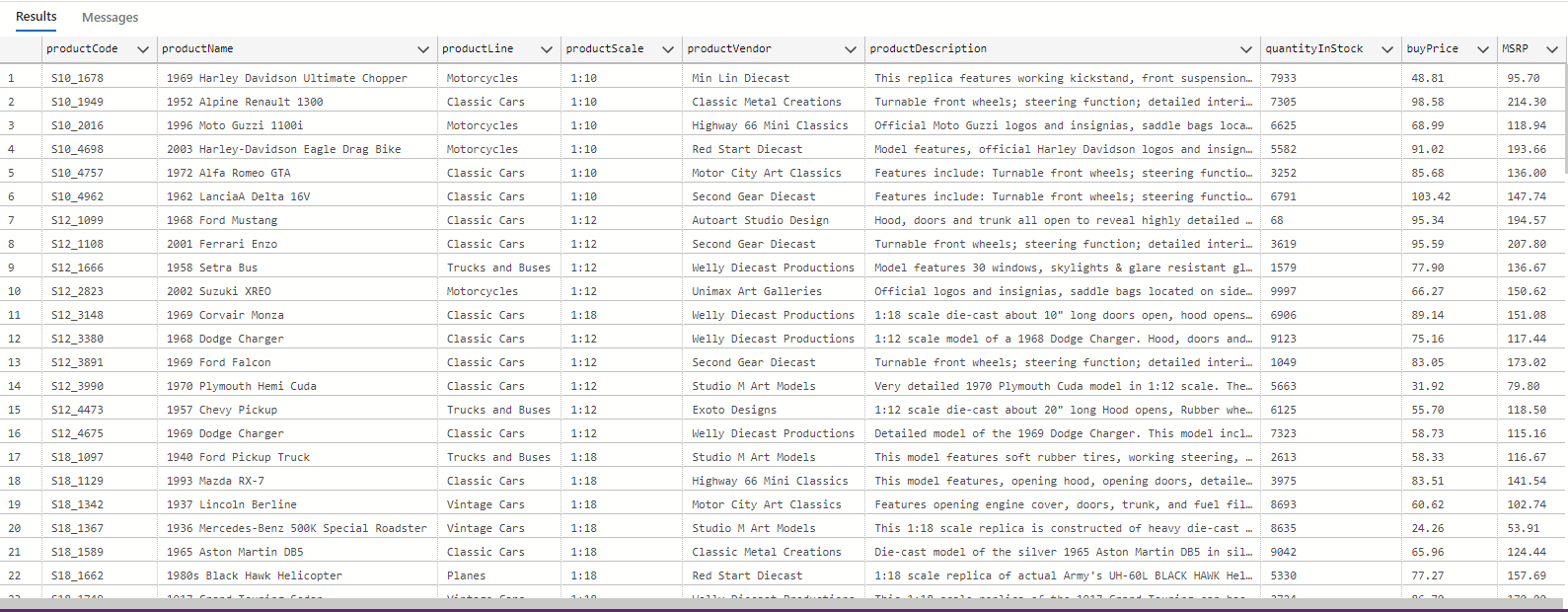
A screenshot of a computer

AI-generated content may be incorrect.

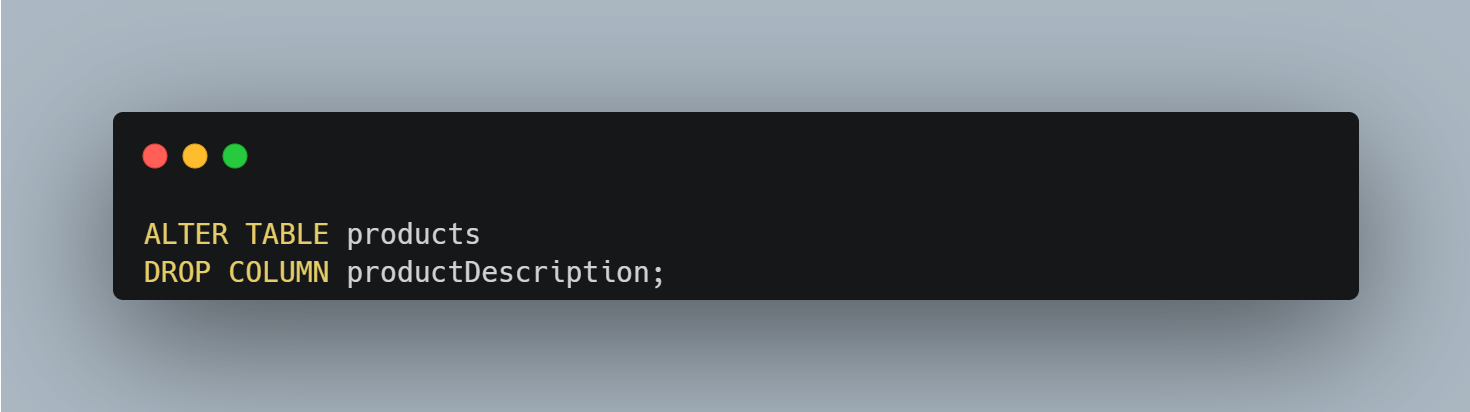
Bu yerdan oxirgi ikki column olib tashlanadi.



Bizning Products tablemiz:



Bu yerdan faqat productDescription table olib tashlaniladi:



Bizning Datasetimiz analiz qilishga tayyor!

E’tiborimizni sotuvni yahshilashga muammolarni aniqlab hal qilishga qaratamiz.

Savollar sone 12ta:

1. Daromadning o’sishi qanday?
2. Eng ko’p sotilgan va eng kam sotilgan maxsulotlar qaysilar?
3. Qaysi oyda bizga ko’proq daromad keladi?
4. Qaysi davlatdan bizga ko’proq daromad kelayabdi?
5. Xodimlarning savdo ko’rsatgichlari qanday?
6. Ombordagi maxsulotlar statistikasi qanday?
7. Rad etilgan mahsulotlar statistikasi qanday?
8. Qaysi mijozlar eng ko‘p buyurtma berishadi?
9. Qaysi mahsulot liniyasi eng ko‘p daromad keltirgan?
10. Qaysi mijozlar lar qanchaga qarzdor?
11. Nechta mijozlar 3 oy ichida buyurtma bermagan?
12. Qaysi mahsulot yetkazib berilishi (shippedDate) eng ko‘p kechikishga uchraydi?

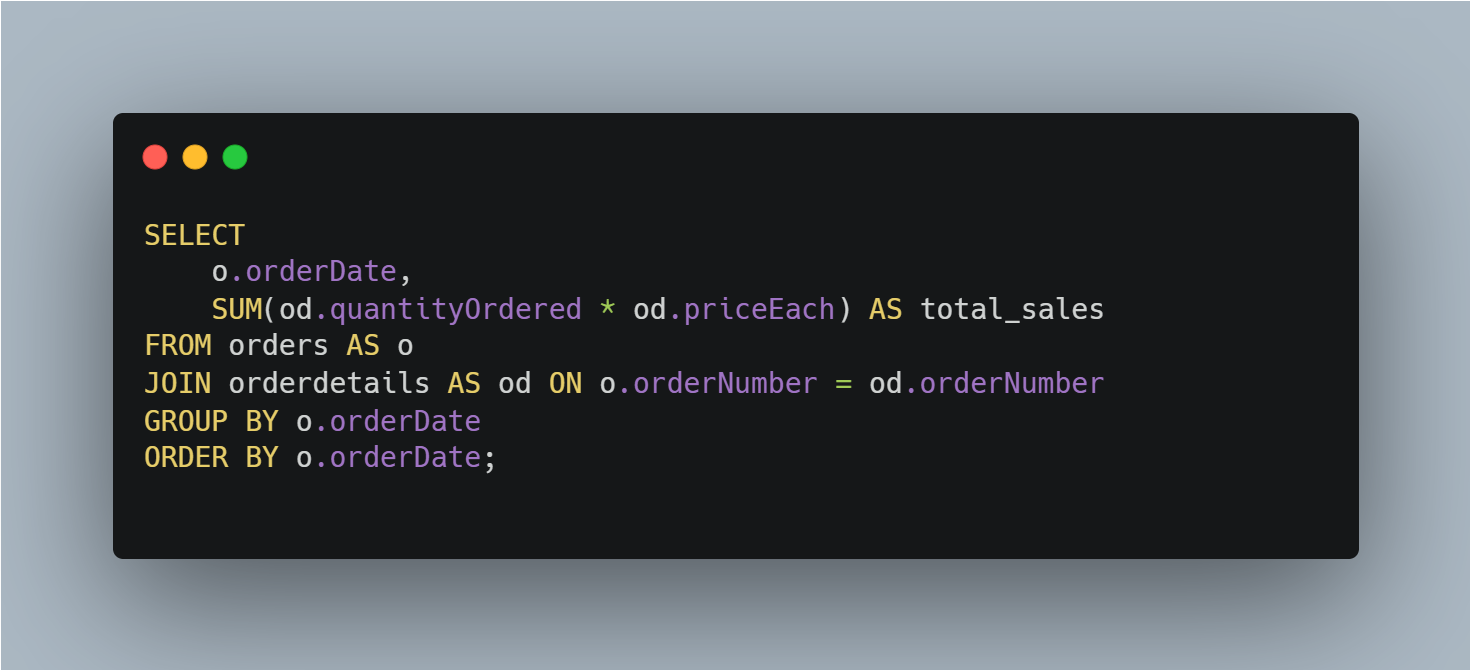
9. Laboratoriya ishi (Ma’lumotlar bazasi o‘qituvchi tomonidan taqdim etiladi):

Ushbu bo’limda biz sotuvni yahshilashga qaratilgan savollarga javob berib boramiz.

1. Savol:

Daromadning o’sishi qanday?

MySQL query yordamida kerak bo’ladigan ma’lumotni export qilamiz:



Power Bi :

A line graph with a blue line

AI-generated content may be incorrect.

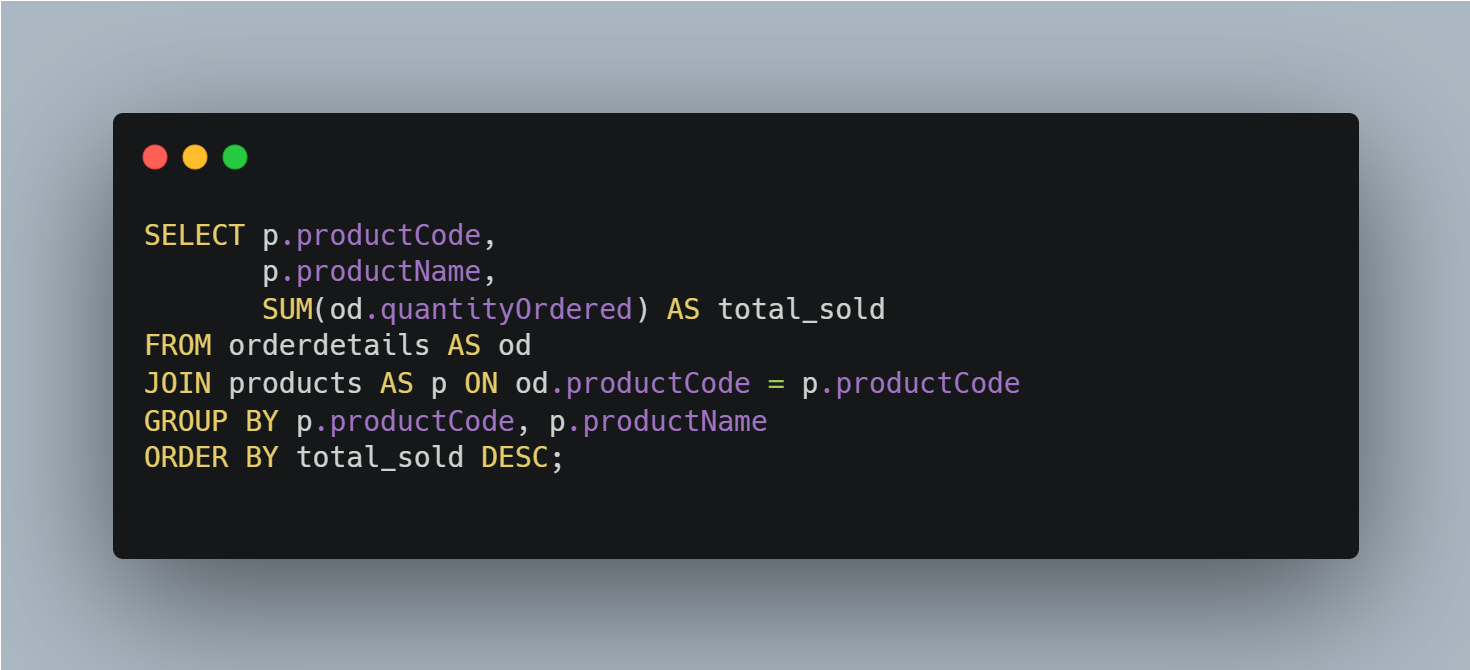
Bu yerda 3 yil mobaynida topilgan daromad ko’rsatilgan. 2004 yil pikka chiqgan. 2005ga kelib tushib ketgan. Birinchi bo’lib shu savolni o’rtaga qo’yganimni sababi sotuvni yahshilash uchun qaratilgan ishni boshlash uchun o’zi Avtomobil savdosi qanday ketayotganini ko’rishimiz kerak edi. 2005-yilda daromad keskin tushib ketgan.

Tavsiya qilamanki sotuvni jarayonini yanada chuqurroq o’rganish kerak. Chegirmalar e’lon qilish kerak. Marketingga katta sarmoya kiritish kerak.

2-Savol:

Eng ko’p sotilgan va eng kam sotilgan maxsulotlar qaysilar?

MySQL query yordamida kerak bo’ladigan ma’lumotni birlashtirib olamiz:



Ushbu savol mahsulotlar kesimida sotuv hajmini solishtirish eng ommabop tovarlarni topish, qaysilar kam ketayotganini aniqlashga yordam beradi.

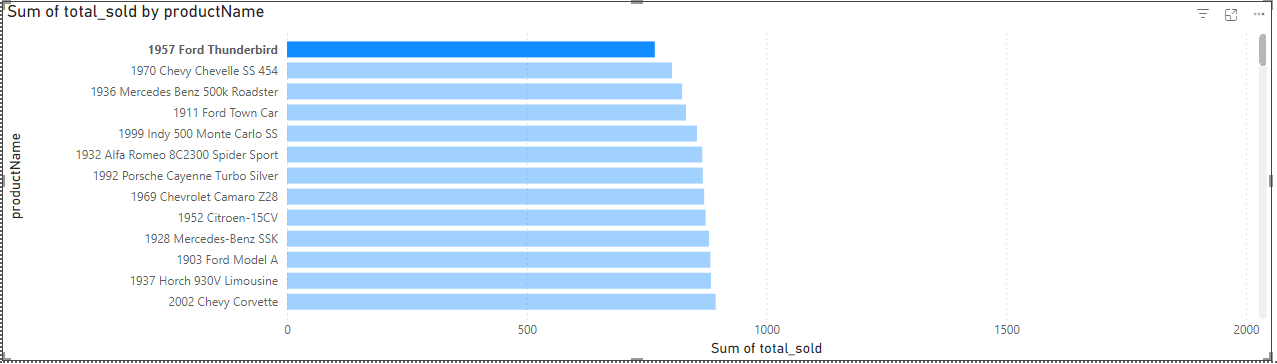
Natijani Powr Bi ga beramiz va Analiz qila boshlaymiz:

A screenshot of a computer

AI-generated content may be incorrect.

Ushbu diagrammada eng ko’p sotilgan product ko’rsatilgan. Satuvda “1992 ferrari 360 red” eng ko’p sotilgan.

Tavsiyam Zahirani ushbu model bilan ta’minlash kerak. Reklamaga urg’u berish kerak.



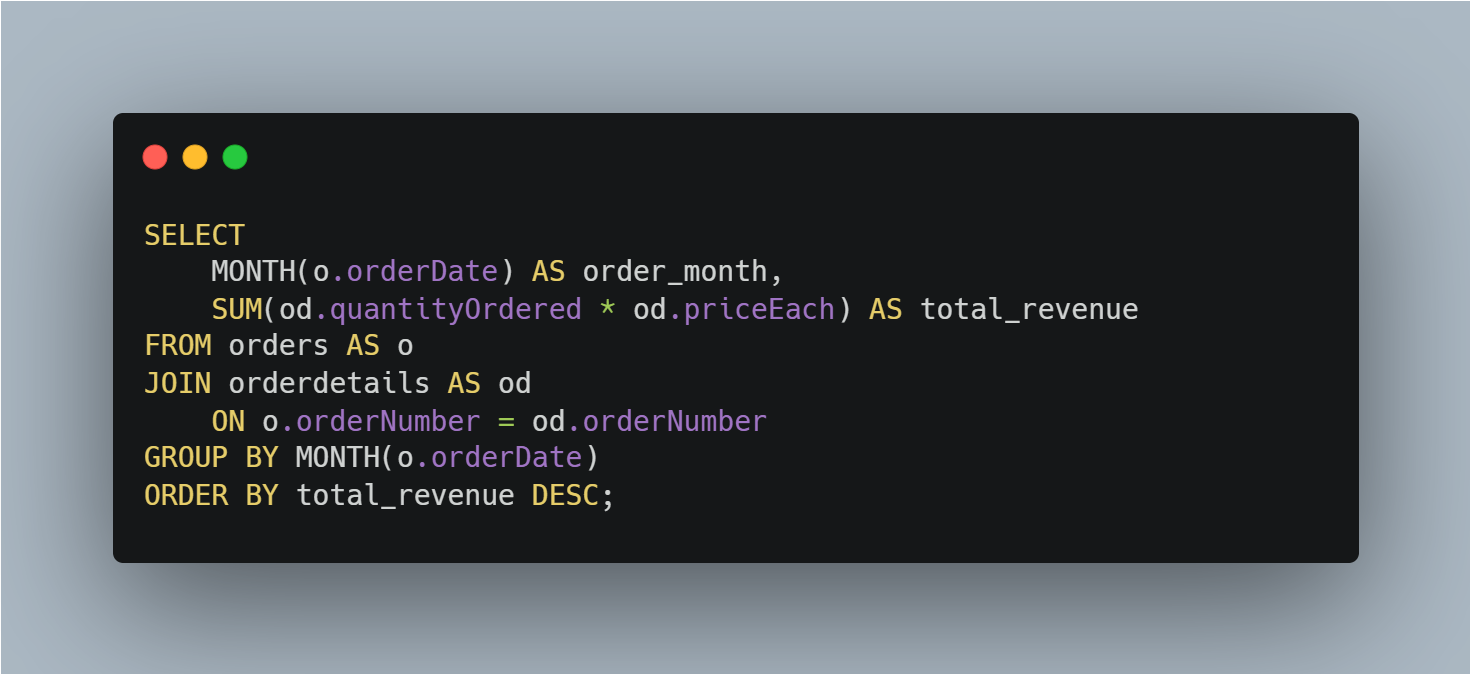
Ushbu diagrammada eng kam sotilgan product ko’rsatilgan. Sotuvda “1957 Ford Thunderbird” eng kam sotilgan.

Narxlarni qayta o’rganib pasaytirish kerak. Skidka e’lon qilish kerak. Assortimentni kamaytirish kerak. Ushbu kam sotilgan modellalrning narxlari, dizayni, sifati bo’yicha mijozlar fikrini o’rganish kerak.

3-Savol

Qaysi oyda bizga ko’proq daromad keladi?

MySQL yordamida keraklik tablelarni birlashtirib olamiz:



Bu savol daromadni oy kesimida ko‘rish, ayniqsa bir necha yil oralig‘ida mavsum bo‘yicha eng yaxshi va eng past natijalarga ega oylarni bilish imkonini beradi.

A line graph with numbers and text

AI-generated content may be incorrect.

Ushbu Line chartda Noyabr oyida eng ko’p daromad bo’lishi ko’rsatilgan, daromad keskin ko’tarilgan. Eng kami Iyunda.

Iyyun oyida aksiyalar e’lon qilish kerak. Noyabr oyiga yaqinlashganda tayyorgarlikni boshlash kerak. Xodimlar va logistika yo’llarini ko’paytirish kertak. Xizmat sifatini yahshilash kerak. Zahirani ko’rib chiqib kam qolgan maxsulotlarni to’ldirish kerak.

4-savol:

Qaysi davlat bizga ko’proq daromad keltirayabdi:

MySql query:

A screenshot of a computer program

AI-generated content may be incorrect.

Bu savol Davlat bo‘yicha savdo hajmini bilish, yahshi daromad keltirayotgan davlatlarda faoliyatni kengaytirish, past bo‘lgan hududlarda esa sabablarni aniqlashimiz kerakligini bildiradi.

Power Bi:



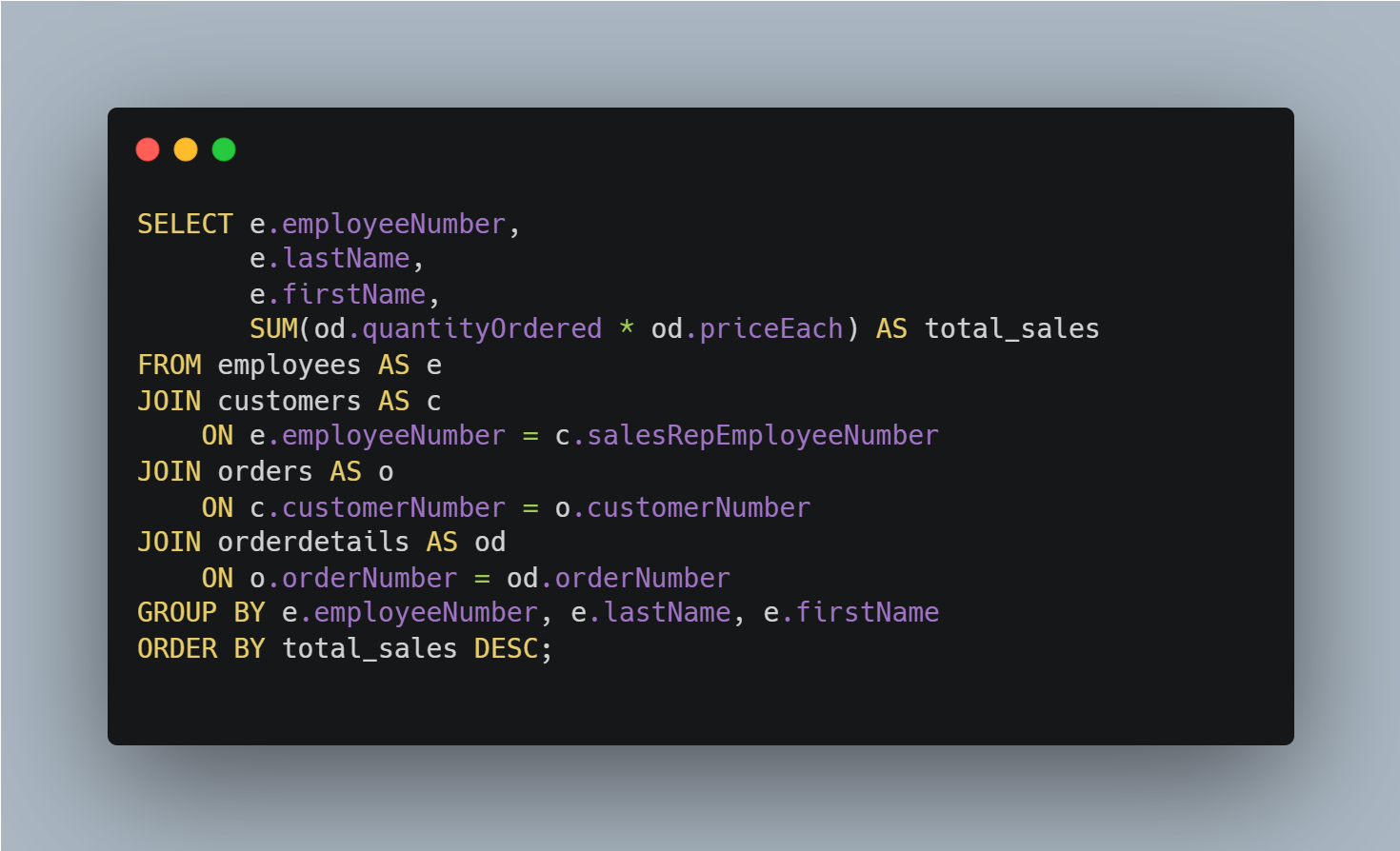
Column chart bizga USA dan eng ko’p daromad kelayotganini ko’rsatayabdi.

Tavsiyalar Eng yuqori daromad berayotgan USA bozorida marketingni kuchaytirib filiallarni ko’paytirish kerak. Kam daromad berayotgan davlatlarda valyuta, logistika yoki bojxonaga oid muammolarni o‘rganish kerak. Mahalliy e’tibor jalb qilish uchun mos marketing strategiyasi ishlab chiqish kerak.

5-savol

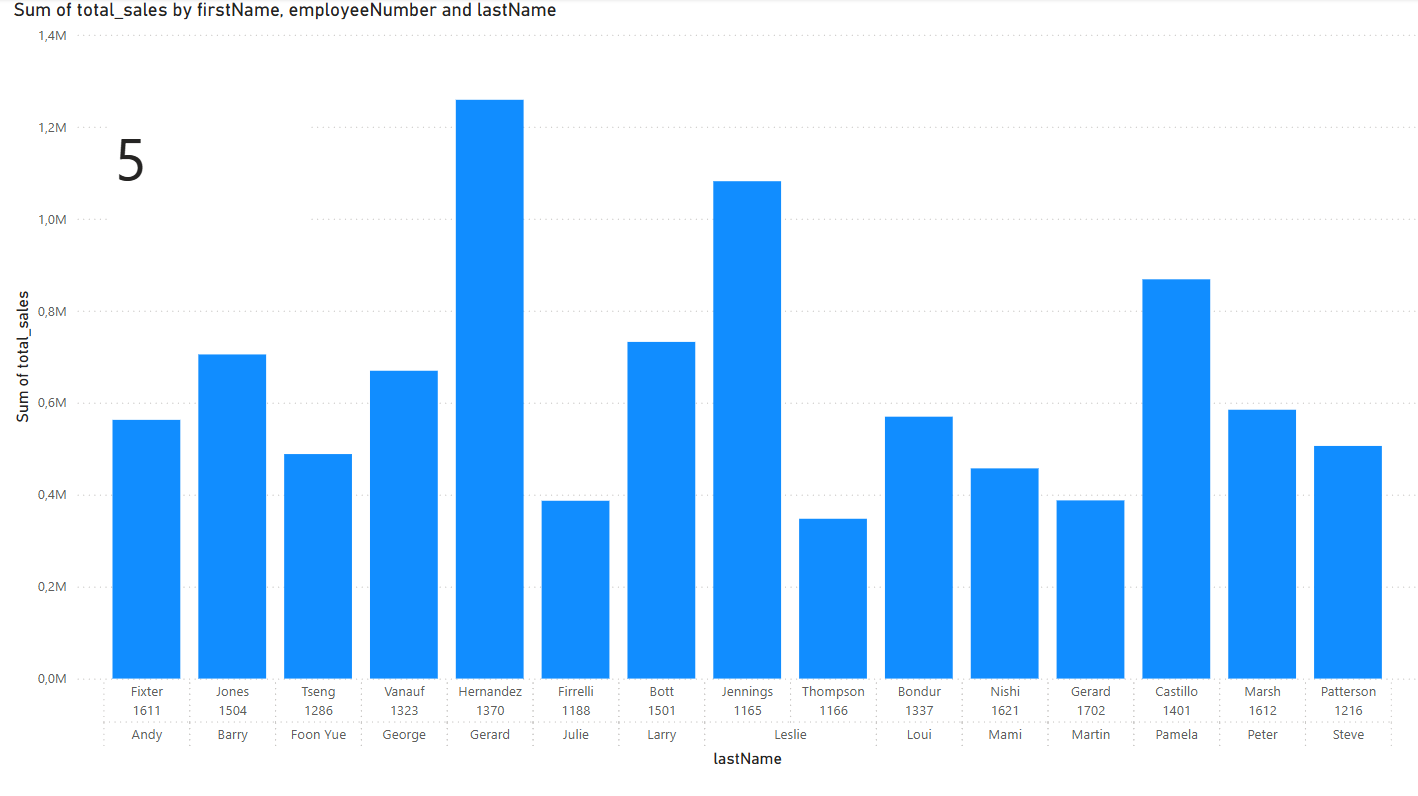
Xodimlarnign sotuv ko’rsatgichi qanday?

MySQL query:



Qaysi xodim (sales representative) qancha sotuv qilganini bilish — mehnat samaradorligi va hodimlarga turtki (motivatsiya) berishda juda muhim.

Power Bi:



Ushbu Colmn chartda ko’rsatilganidek eng yuqori sotuv ko’rsatgichiga ega xodim Gerard Hernandez.

Tavsiyalar:

Har oyda Sotuvning samaradorligini yahshilash maqsadida hodimlarni Rag’batlantirish kerak. Past ko’rsatgichdagi xodimlarni o’qitish treninglar o’tkazish kerak. Samaradorlikni kuzatib borish kerak.

6- Savol

Ombordagi mahsulotlar statistikasi qanday?

MySQL query:



Bu savol Inventory darajasni bilish, qanday tovar zaxirada ortib ketgan, qaysi biri kam qolgan? logistika va ta’minot zanjirini boshqarishda asqotadi.

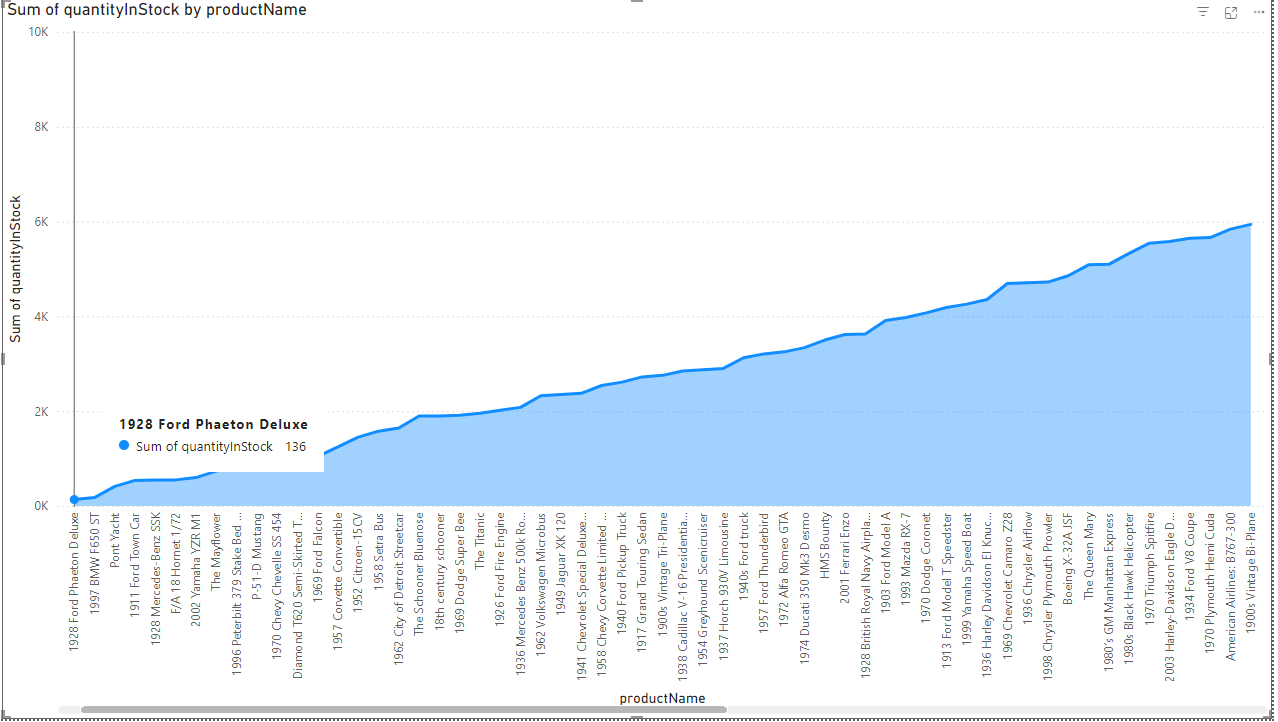
Power Bi:

A blue and white graph

AI-generated content may be incorrect.

“2002 Suzuki XREO” avtomobili 9997ta qolgan.

Eng kam qolgan mahsulot esa:



Area chartda tasvirlanganidek omborda eng kam qolgan maxsulot “1928 Ford Pheaton Deluxe” (136ta qolgan).

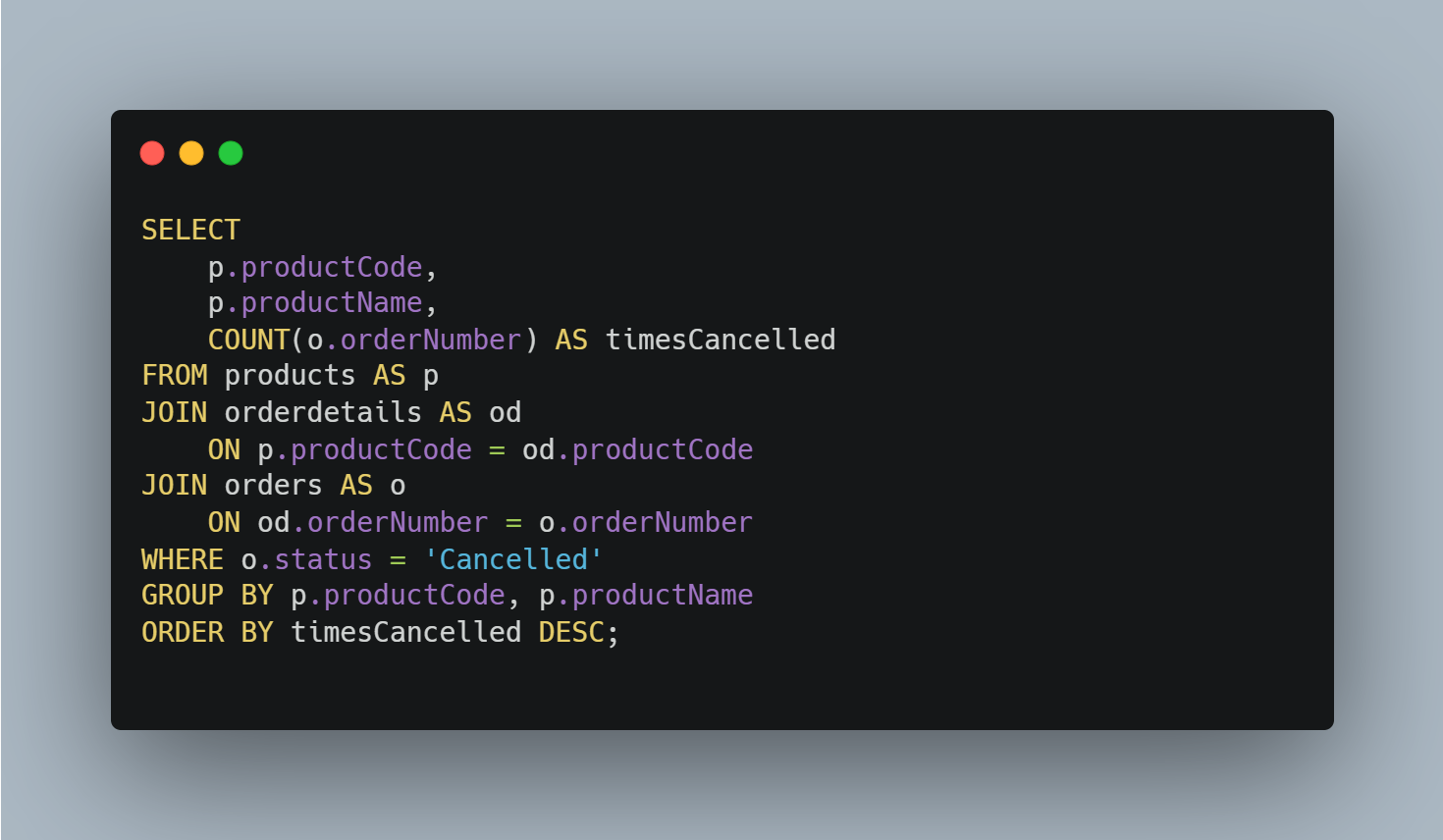
Tavsiyalar:

Eng ko‘p sotiladigan texnikalar zaxirasini yetarli darajada ushlab turish kerak. Yeg’ilib ko’payib ketgan ya’ni sotilmayotgan mahsulotlarni tezroq sotilishi uchun chegirmalar va promo aksiya o’ylab topish kerak.

7-savol

Rad etilgan mahsulotlar statistikasi

MySQL so’rov :



Ushbu savol buyurtma rad etilgan buyurtmadagi mahsulotlar nimaligini bilish, rad etishga sabab bo’lgan sifat, narx, ombordagi muammolarni tekshirish zarur.

Power Bi:



Eng ko’p rad etilgan mahsulot “1940 Delivery Sedan” va “1999 yamaha speed boat”.

Unchalik ko’p rad etilmagan bo’lsa ham daromadga ta’siri bor.

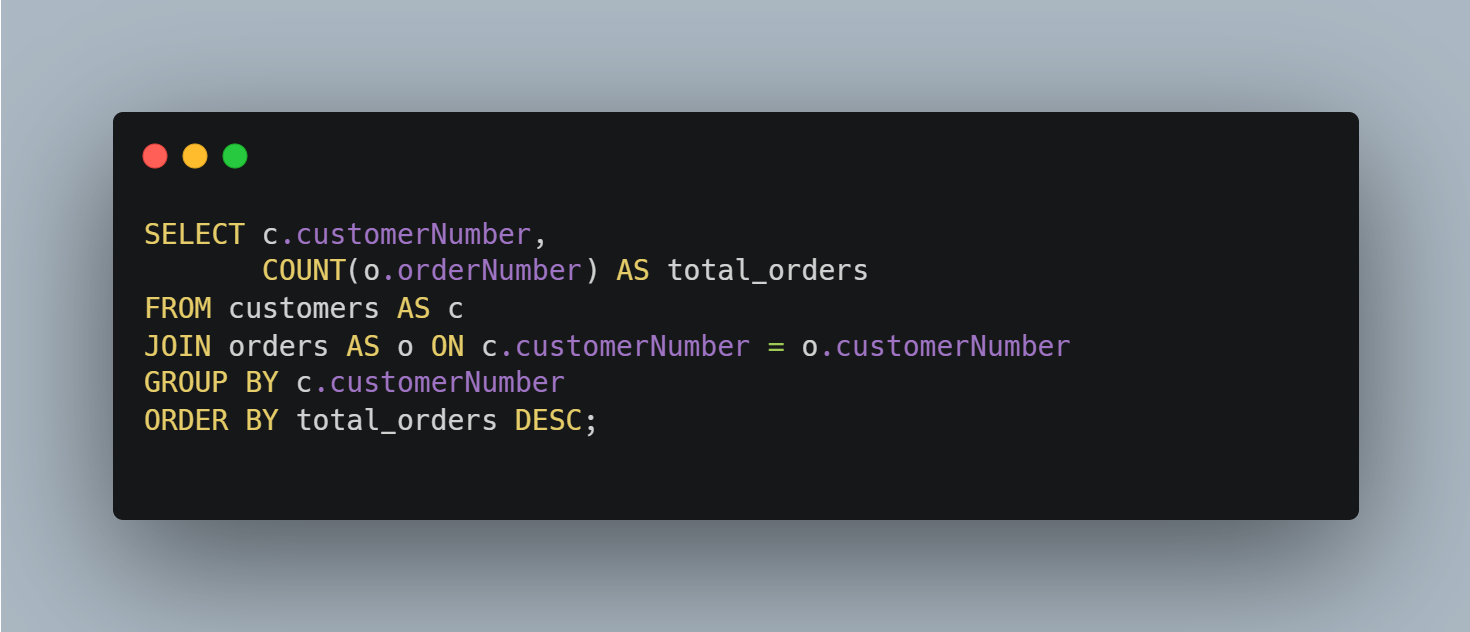
Tavsiyam:

Ushbu mahsulotlar bo’yicha ishlab chiqarish yoki sifat masalasini ko‘rib chiqish lozim. Ba’zan narx balandligi yoki kech yetkazish ham rad etishga sabab bo‘ladi. Rad etish sabablarini aniq savolnoma (feedback) orqali oydinlashtirilishi kerak.

8- savol:

Qaysi mijozlar eng ko‘p buyurtma berishadi?

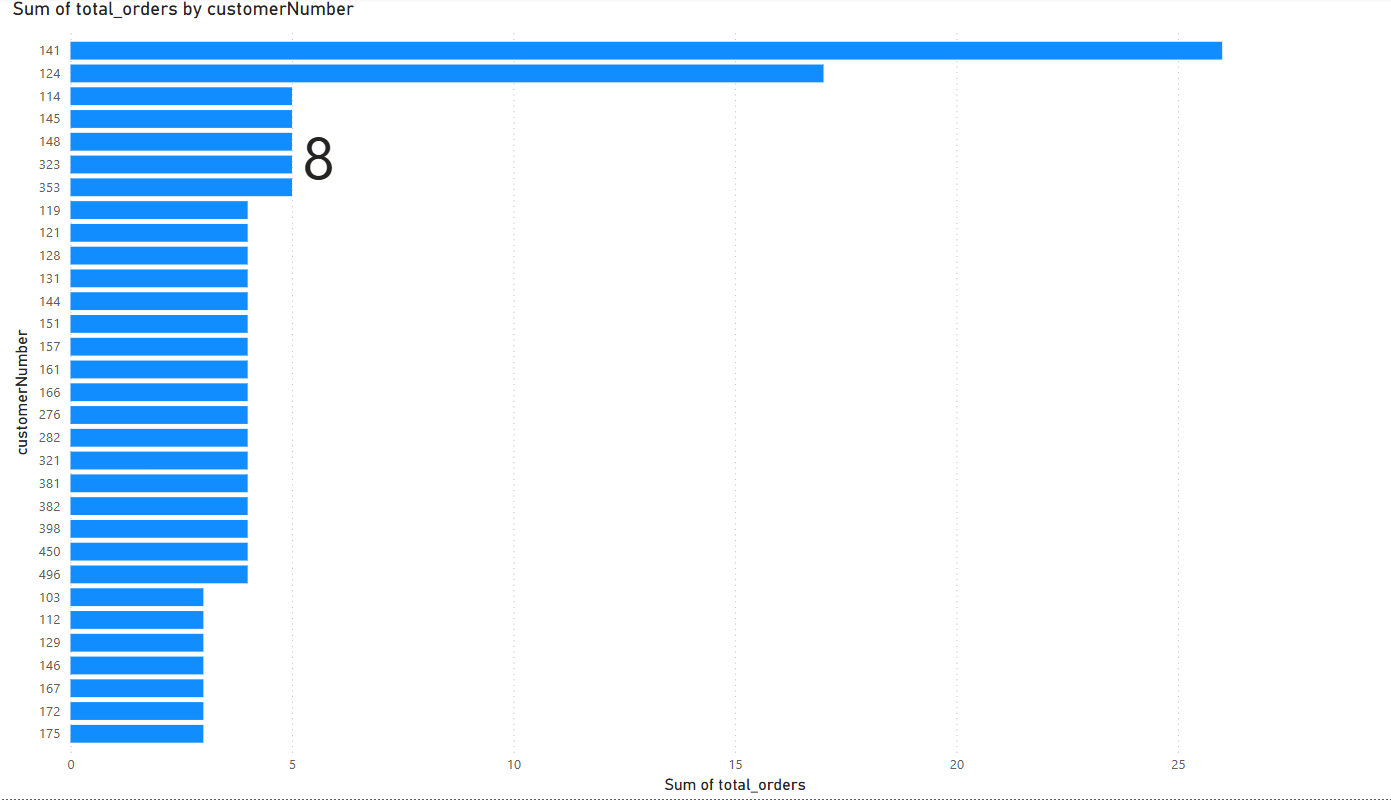
Birlashtirish uchun Query:



Biz anonimlikni saqlash maqsadida mijozlarning ismlarini olib tashlagan edik, shuning uchun CustomerNumber dan foydalandik.

Ushbu savol mijozlarning umumiy buyurtma soni yoki summasini tahlil qilish — asosiy sodiq mijozlar (loyal customers) kim ekanini bilishga yordam beradi.

Power Bi:



141- raqamli customer eng ko’p buyurtma bergan.

Tavsiyalarim:

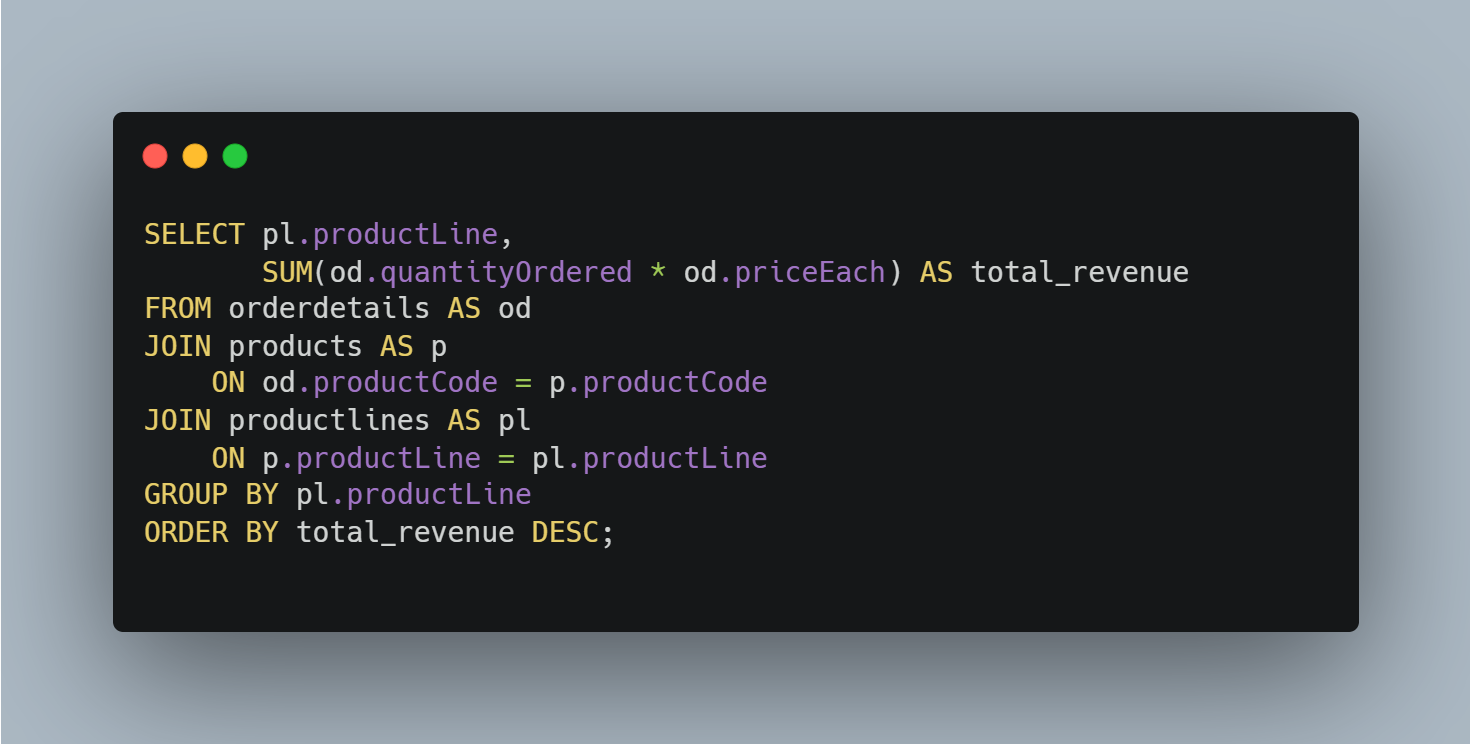
Sodiq mijozlar bilan ishlash kerak, ularga maxsus bonus yoki VIP dastur taklif qilish kerak. Qo’llab quvvatlash kerak.

Ushbu ko‘p buyurtma qilgan mijozlardan referal (tanishlarini taklif qilish) dastur orqali yangi mijozlar jalb qilish kerak.

9- savol

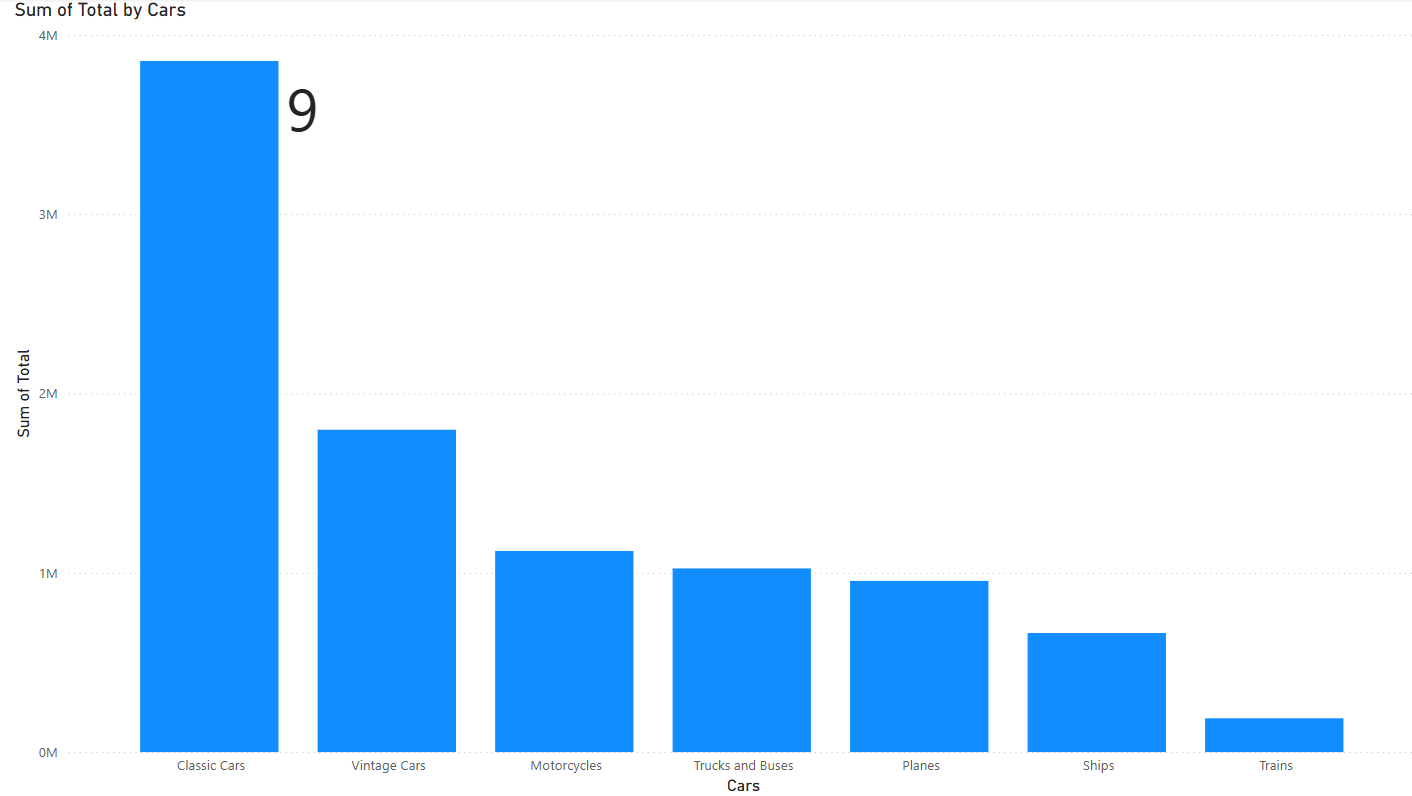
Qaysi mahsulot liniyasi ko‘p daromad keltirgan?

Query:



Mahsulot liniyalari bo‘yicha tahlil kompaniya qaysi toifaga ko‘proq e’tibor berishi kerakligini ko‘rsatadi.

Power Bi:



Eng katta oqim bizga Classic car turkumidan keladi. Foyda 3, 854mlnga yaqin.

Tavsiyalar:

Eng daromadli liniya Classic Cars turkumiga sarmoya kiritish kerak, mahsulot sifatini, marketingini kuchaytirish kerak. Kam daromad kelayotgan liniyalarni kam daromad kelayotganlik sababini o’rganish kerak, ushbuturkumga oid raqiblarni o’rganish kerak. Rebrand Redizayn xaqida o’ylab ko’rish kerak.

10-savol:

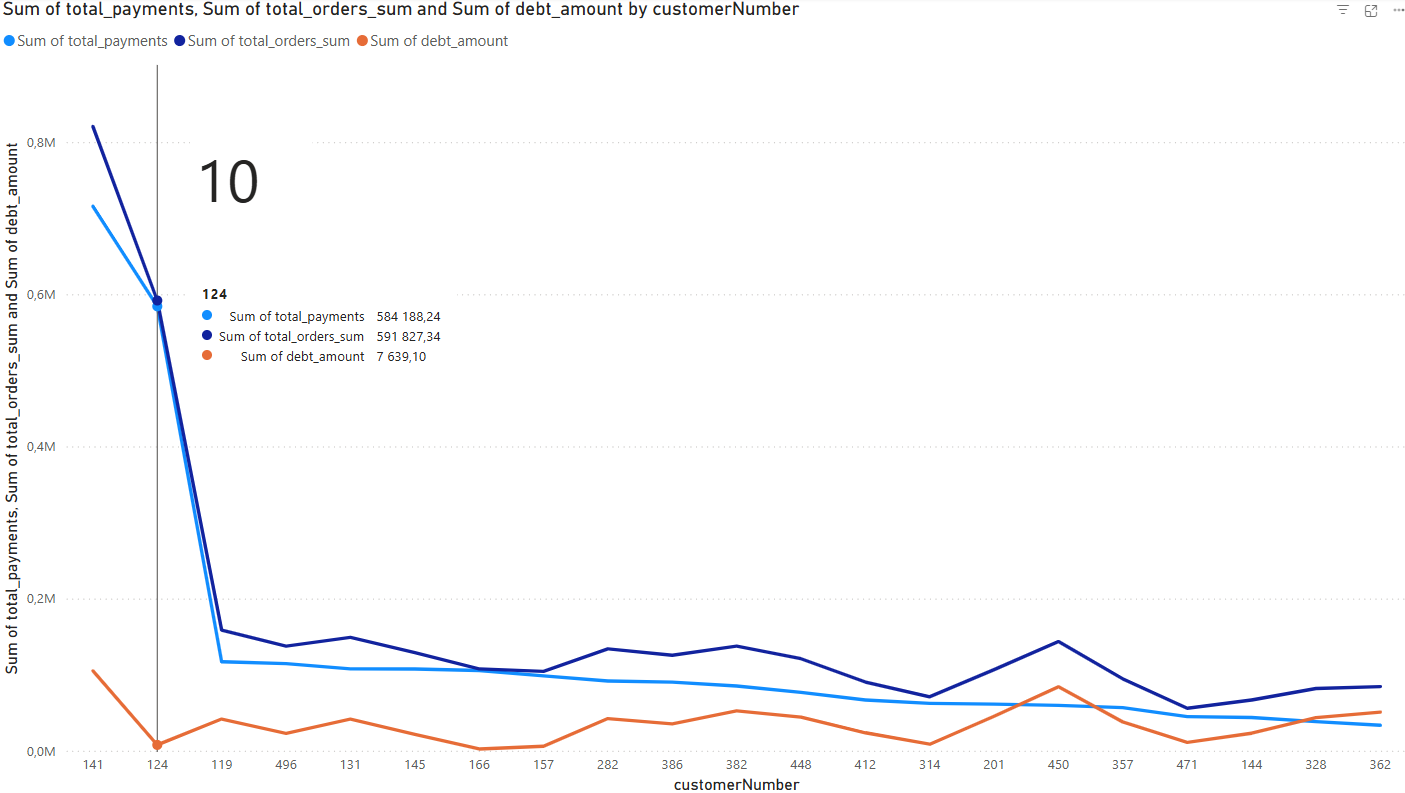
Qaysi mijozlar lar qanchaga qarzdor?

Query:



Ushbu savol mijozlarning to‘lovlari buyurtma summasiga nisbatan kechikib ketishi, qarzdorlik (overdue) kechikmalarini bilish naqd oqimini saqlashda juda muhim.

Power Bi:



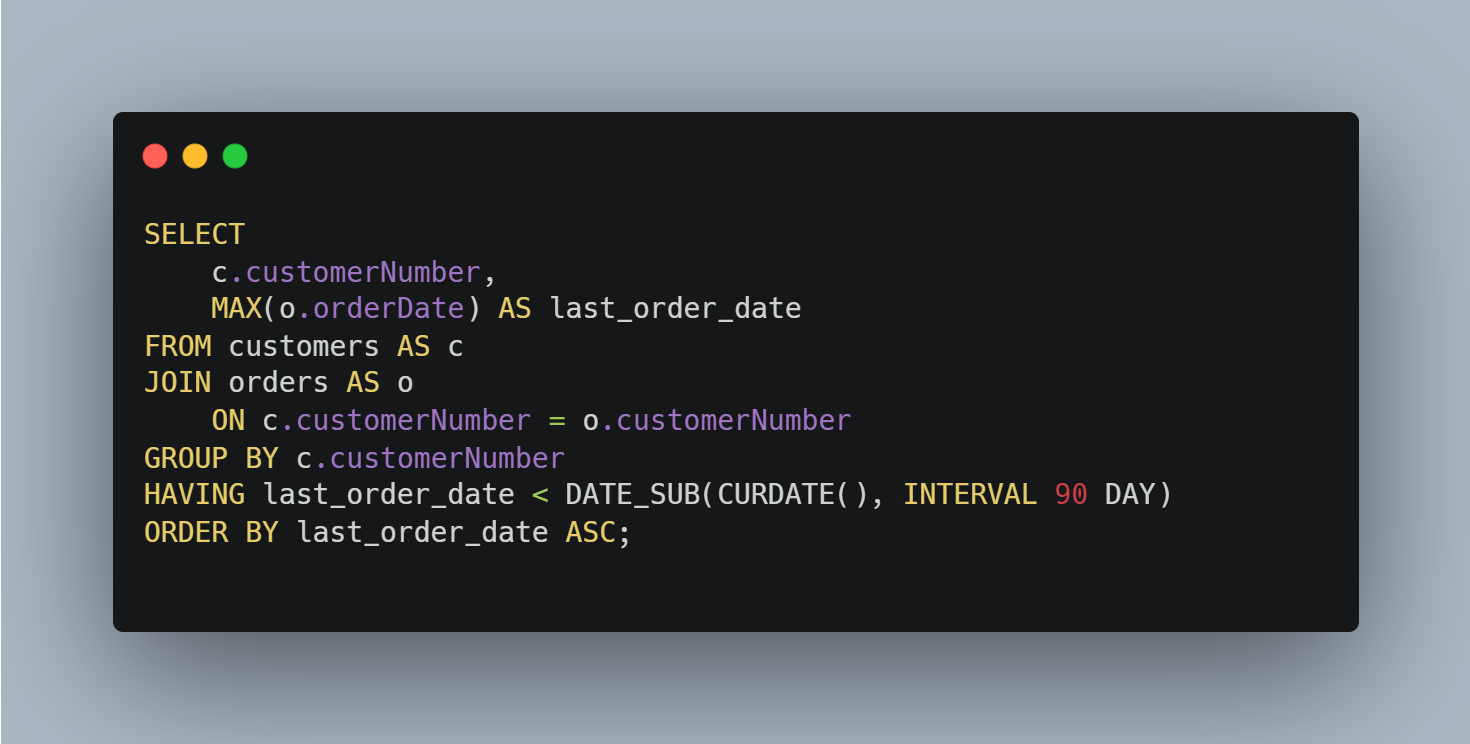
141-mijozning qarzdorligi 104,950.56ni tashkil qiladi.

Ushbu yirik qarzdorlarga bo‘lib to‘lash muddatlarini taklif qilish orqali muzokara yuritish kerak. Qarzdor mijozlarga tez-tez xabar berish kerak. Qarzini to’lashligi uchun vaqt belgilab borish kerak. Risklar bilan ishlash kerak. Ya’ni mijoz tomonidan yangi buyurtma berilganida qarzdor mijozlar bilan “oldindan to‘lov” siyosatini qo‘llash kerak.

11-savol

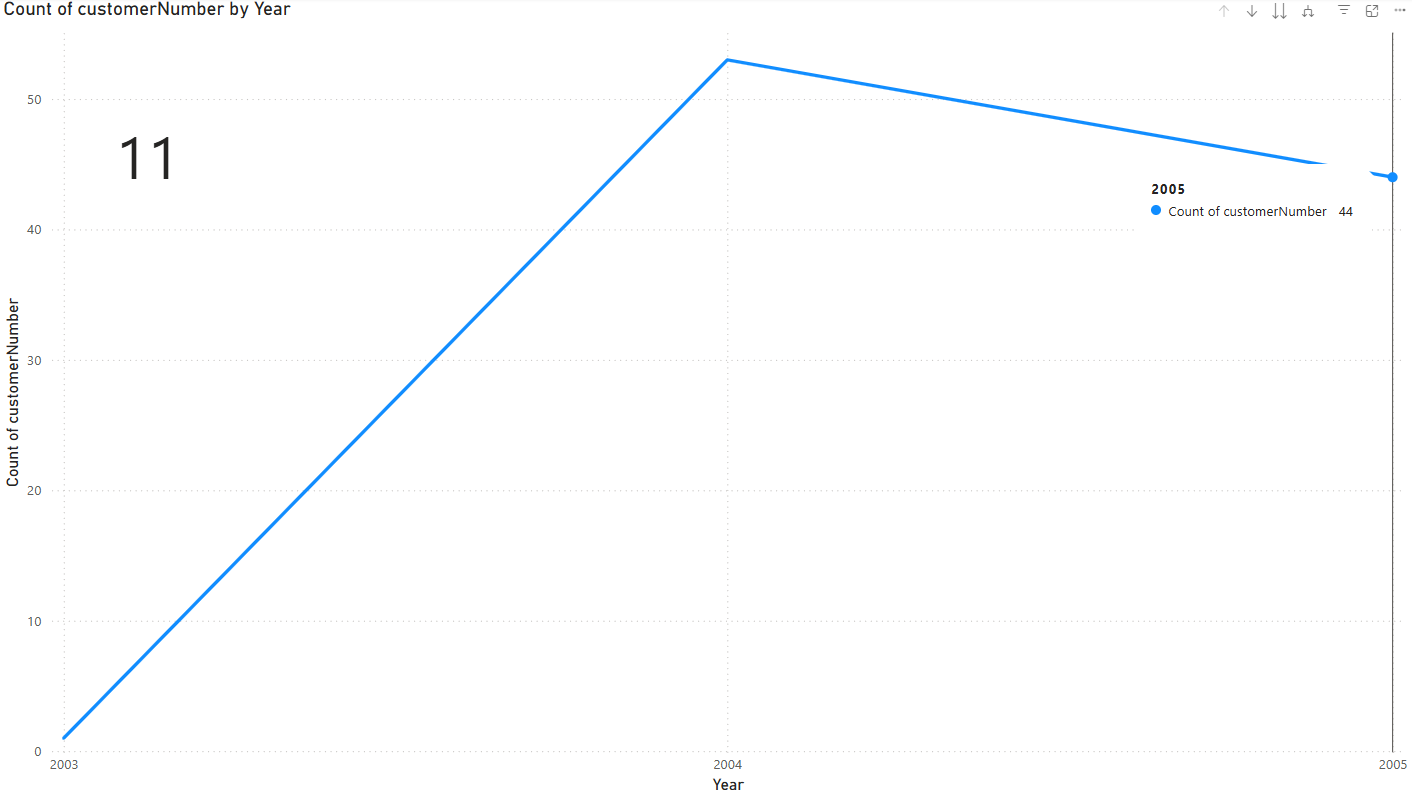
Nechta mijozlar 3 oy ichida buyurtma bermagan?

Mysql query:



Ushbu savol mijozlarni yo‘qotish (churn) xavfi bilan ishlashga yordam beradi. So‘nggi 3 oy ichida aktiv bo‘lmagan mijozlar kim? Ular bilan qayta aloqa o‘rnatish orqali savdoni oshirish mumkin.

Power Bi:



2004 yilda 53 mijoz qaytib buyurtma bermagan. 2005-yilda esa 44ta mijoz.

Tavsiyalar:

CRM sistemasini o’rnatish kerak. Qayta jalb qilish kerak. Ushbu mijozlar uchun individual chegirma, promo kod yoki qo‘ng‘iroq orqali taklif berish kerak.

Churn sababini o’rganish kerak. Xizmat sifati, narx yoki raqobatchi takliflari sabab bo‘lishi mumkin. Mijozlarni ushlab qolish uchun ball, mukofot yoki cashback tizimini o’rnatish kerak.

12-savol

Qaysi mahsulot yetkazib berilishi (shippedDate) eng ko‘p kechikishga uchraydi?

Query:



Buyurtma sanasi (orderDate) va jo‘natish sanasi (shippedDate) o‘rtasidagi farq qanchalik katta? Bu savol logistikadagi muammolarni aniqlashda va mijozlarning raqobatlarga kechikish sababli ketib qolishining oldini olishda yordam beradi.

Power Bi :

A graph with a line going up

AI-generated content may be incorrect.

“1970 Triumph Spitfire” eng ko’p kechiktirishga uchraydi.

Tavsiyalar:

Kechikish ko‘p bo‘lgan mahsulotlarning sababini topish kerak. Zaxira yo‘qligi, etkazib beruvchi kechiktirgani, tayyorlash kechikishi sabab bo’lishi mumkin. Tashqi servis bilan hamkorlik qilinsa, shartnomalarni qayta ko‘rib chiqish kerak, tezkor yetkazish tariflarini ko‘rib chiqish kerak. Eng ko‘p kechikadigan mahsulotlar talabini oldindan hisoblash, yetkazib berishni optimallashtirish kerak.

10. Ma’lumotlar bo'yicha mutaxassislar duch keladigan turli rollar, mas'uliyat va muammolarni tushuntiring.

Ma’lumotlar bo‘yicha mutaxassislar turli yo‘nalish va lavozimlarda ishlaydi. Ularning asosiy vazifasi tashkilotdagi ma’lumotlarni samarali yig‘ish, saqlash, tahlil qilish va undan foydali xulosalar chiqarishdan iborat. Quyida ularning asosiy rollari va mas’uliyatlari ko‘rsatilgan: Ma’lumotlar muhandisi (Data Engineer) Ma’lumotlarni olish, tozalash, qayta ishlash va markazlashgan holda saqlash infratuzilmasini yaratadi. Ularning vazifalari quyidagilarni o‘z ichiga oladi: ETL/ELT jarayonlarini sozlash Ma’lumotlar ombori yoki Data Lake kabi manbalarni yaratish va boshqarish Ma’lumotlarni real vaqt rejimida yoki paket shaklida uzatish Bulut platformalarini (AWS, Azure, GCP) sozlash va optimallashtirish Ushbu rollarda muhandislar quyidagi muammolarga duch kelishi mumkin: heterojen manbalar katta hajmdagi ma’lumot skhema o‘zgarishlari va ish faoliyatini optimallashtirish Ma’lumotlar tahlilchisi (Data Analyst) Ma’lumotlar tahlilchisi biznes savollariga javob berish uchun mavjud ma’lumotlarni so‘rovlar statistika va vizualizatsiya yordamida o‘rganadi.

Ularning vazifalari: ma’lumotlarni tozalash va transformatsiya qilish KPI aniqlash va hisobot tayyorlash Excel Power BI yoki Tableau vositalarida dashbordlar yaratish va rahbariyatga xulosa taqdim qilish Muammolar iflos ma’lumotlar noaniq biznes talablari mos kelmaydigan jadvallar va tezkorlikni talab qiluvchi ishlar bilan bog‘liq bo‘ladi. Ma’lumotlar olimi (Data Scientist) Machine learning statistik model yoki prognozlash metodlari yordamida murakkab savollarga yechim izlaydi. Vazifalari model ishlab chiqish patternlar topish va model natijalarini kuzatish Ma’lumotlar yetarli bo‘lmasligi modelni tushuntirish va muntazam yangilash kabi muammolarga olib kelishi mumkin. Biznes intellekt dasturchisi yoki Dev (BI Developer) BI platformalarini o‘rnatish va murakkab hisobotlar yaratish bilan shug‘ullanadi. Ularning mas’uliyatlari ma’lumotlar modelini loyihalash OLAP kublarni optimallashtirish va BI servislarini sozlashdan iborat. Ular birlashtirilmagan ma’lumot manbalari dasturiy moslik va foydalanuvchi talablari kabi muammolar bilan duch keladi (Smith et al., 2020).

11. Ma'lumotlar muvofiqligini ta'minlash uchun ma'lumotlar mutaxassislari tomonidan qo'llaniladigan turli strategiyalarni ko'rib chiqing

Ma'lumotlar izchilligini ta'minlash turli tizimlar va manbalarda ma'lumotlarning ma'nosi, formati va talqinida bir xillikni saqlashni anglatadi. Quyida ma'lumotlar bo'yicha mutaxassislar (ma'lumotlar muhandislari, ma'lumotlar tahlilchilari, ma'lumotlar olimlari, ma'lumotlar arxitektorlari va boshqalar) tomonidan qo'llaniladigan asosiy strategiyalar mavjud.

1. Ma'lumotlar bazasini to'g'ri loyihalash

Normalizatsiya va cheklovlar

Yaxshi ishlab chiqilgan ma'lumotlar bazasi ortiqcha ma'lumotlarni kamaytirish va nomuvofiqliklarning oldini olish uchun normallashtirishni (1NF, 2NF, 3NF va boshqalar) o'z ichiga oladi (Inmon, 2005; Kimball va boshq., 2013). PRIMARY KEY, FOREIGN KEY, UNIQUE va CHECK kabi cheklovlar ma'lumotlarning belgilangan qoidalar doirasida qolishini ta'minlaydi (Loshin, 2009).

Ma'lumotlarni boshqarishning asosiy usuli (MDM)

MDM barcha manbalar bo'ylab asosiy ob'ektlarning (masalan, mijozlar yozuvlari) yagona, aniq versiyasini saqlaydigan "oltin rekord" tushunchasidan foydalanadi (Loshin, 2009). Muhim atributlarni (nomlar, manzillar, identifikatorlar) markazlashtirib, MDM bir nechta tizimlar bo'ylab ziddiyatli ma'lumotlar xavfini kamaytiradi.

2. ETL/ELT jarayonlarida tozalash va tekshirish

Extract-Transform-Load (ETL) yoki Extract-Load-Transform (ELT) quvurlari heterojen manbalardan olingan ma'lumotlarni qayta ishlaydi (Hyde, 2021). Transformatsiya bosqichida:

Ma'lumotlarni tozalash iflos, takroriy yoki to'liq bo'lmagan yozuvlarni olib tashlaydi yoki tuzatadi.

Ma'lumotlarni standartlashtirish izchil ma'lumotlar turlarini, formatlarini (masalan, sana formatlari) va kodlashni ta'minlaydi.

De-duplikatsiya bir xil ob'ektga tegishli takrorlangan yozuvlarni birlashtiradi.

Har bir bosqichda tasdiqlash qoidalari (masalan, "salbiy raqamlar yo'q", "yaroqli sanalar oralig'i", "mos keladigan xorijiy kalitlar") ma'lumotlar omboriga yoki ma'lumotlar ko'liga yuklashdan oldin izchil ma'lumotlarni saqlashga yordam beradi.

3. Ma'lumotlarni boshqarish va rollarni belgilash

Ma'lumotlarni boshqarish tizimi ma'lumotlarga qanday egalik qilish, undan foydalanish va nazorat qilinishini tavsiflaydi (DAMA International, 2017). Asosiy tarkibiy qismlarga quyidagilar kiradi:

Rollarni aniq belgilash: ma'lumotlar egasi, ma'lumotlar boshqaruvchisi va ma'lumotlar iste'molchisi belgilangan mas'uliyatga ega.

Ma'lumotlar katalogi: har bir atributning formatini, domenini va manfaatdor tomonlarini hujjatlashtiradi, bu ma'lumotlarning kelib chiqishini kuzatishni osonlashtiradi.

Ma'lumotlar nasl-nasabi: ma'lumotlarning turli xil o'zgarishlar orqali qanday o'tishini va oxir-oqibat qaerda joylashganligini kuzatib boradi.

Bunday boshqaruv amaliyotlari ma'lumotlarning izchilligi bilan bog'liq muammolarni tezda aniqlash va kerakli guruhlarga etkazishni ta'minlaydi.

4. Real vaqt rejimida monitoring va avtomatlashtirilgan sinov

Haqiqiy vaqtda yoki real vaqtga yaqin monitoring ma'lumotlar hajmining kutilmagan tushishi yoki keskin ko'tarilishini aniqlaydi va ogohlantirishlarni oshiradi. Avtomatlashtirilgan sinov vositalari (masalan, Great Expectations, dbt tests, Deequ) ma'lumotlar sifatini tekshiradi va qiymatlar kutilgan chegaralardan tashqariga tushganda ogohlantirishlarni ishga tushiradi (Hyde, 2021). Ushbu testlarni CI/CD quvurlariga integratsiyalash yangi o'rnatishlar mavjud ma'lumotlar oqimini buzmasligini ta'minlaydi.

5. Ma'lumotlar ombori va ma'lumotlar ko'li arxitekturasi

Ma'lumotlar ombori (DWH) usullari taniqli yondashuvlarga amal qiladi (Inmon, 2005; Kimball va boshq., 2013). DWH toza, mavzuga yo'naltirilgan va tarixiy ma'lumotlarni saqlaydi va shu bilan izchillikni saqlaydi.

Data Lake platformalari bitta markaziy omborda xom yoki yarim tuzilgan ma'lumotlarni joylashtiradi. "Ma'lumotlar botqog'i" ning oldini olish uchun mustahkam katalog va sxemalarni qo'llash qobiliyati juda muhimdir (Databricks, 2022).

Lakehouse arxitekturalari ma'lumotlar omborlari va ma'lumotlar ko'llarining afzalliklarini birlashtirib, ACID tranzaktsiyalari va kengaytiriladigan saqlashdan foydalanadi.

6. Versiya va sxemalarni boshqarish

Sxema o'zgarishlari (masalan, ustunlarni qo'shish, nomini o'zgartirish yoki o'chirish) to'g'ri bajarilmasa, nomuvofiqliklarga olib kelishi mumkin:

Sxema registri (masalan, Avro, Kafka uchun) har bir versiyani kuzatib boradi.

Orqaga/oldinga muvofiqligi eski jarayonlarning yangi maydonlar yoki yangilangan formatlar bilan ishlashini ta'minlaydi.

7. Xulosa

Sog'lom arxitektura (DWH, Data Lake, MDM) va to'g'ri dizayn (normalizatsiya, cheklovlar) izchil ma'lumotlar uchun asos yaratadi.

To'liq tozalash va tekshirish bilan ETL/ELT o'zgarishlari heterojen ma'lumotlar manbalarini birlashtiradi.

Rol va mas'uliyatni, ma'lumotlar katalogini va nasl-nasabini belgilovchi ma'lumotlarni boshqarish tizimi nomuvofiqliklarni aniqlash va hal qilishga yordam beradi.

Monitoring, sinov va CI/CD quvurlari xatolarni aniqlash va tuzatish harakatlarini avtomatlashtiradi.

Sxema evolyutsiyasi strategiyalari vaqt o'tishi bilan tuzilmalar muqarrar ravishda o'zgarganda ma'lumotlarning mustahkamligini ta'minlaydi.

Ma'lumotlarning izchilligini bir martalik vazifa emas, balki doimiy jarayon sifatida ko'rib, tashkilotlar o'zlarining tahliliy, operatsion va strategik maqsadlari uchun aniq, ishonchli ma'lumotlardan foydalanishlari mumkin.

12. Ma’lumotlarga asoslangan madaniyatga axloqni shakllantirishda ma'lumotlar mutaxassislari duch keladigan rol, mas'uliyat va muammolarni tahlil qiling.

Ma’lumotlarga asoslangan madaniyat (Data-Driven Culture) tashkilotning barcha jabhalarida qaror qabul qilish jarayonlarini ma’lumotlar tahlili natijalariga asoslashga qaratilgan yondashuvdir. Bunday madaniyatni shakllantirishda ma’lumotlar mutaxassislari muhim rol o‘ynaydilar, lekin ular bir qator mas’uliyat va muammolarga duch keladilar.

Ma’lumotlar mutaxassislarining roli ma’lumotlarni yig‘ish, qayta ishlash, tahlil qilish va ular asosida xulosalar chiqarishdan iborat. Ular tashkilotga ma’lumotlardan samarali foydalanish strategiyasini ishlab chiqishda, ma’lumotlar sifatini ta’minlashda va ma’lumotlar tahlili natijalarini boshqa xodimlarga yetkazishda yordam beradilar (Provost & Fawcett, 2013). Bundan tashqari, ular ma’lumotlar tahlili bilan bog‘liq axloqiy muammolarni hal qilishda ham ishtirok etadilar.

Ma’lumotlar mutaxassislarining mas’uliyati quyidagilarni o‘z ichiga oladi:

Ma’lumotlarning maxfiyligini ta’minlash: Shaxsiy ma’lumotlarni himoya qilish va ulardan faqat qonuniy maqsadlarda foydalanish (Zuiderveen Borgesius et al., 2018).

Ma’lumotlarning adolatli ishlatilishi: Ma’lumotlar tahlili natijalaridan diskriminatsiya va boshqa adolatsiz amaliyotlar uchun foydalanmaslik.

Ma’lumotlarning to‘g‘riligini ta’minlash: Tahlil natijalarining ishonchliligini ta’minlash uchun ma’lumotlarni to‘g‘ri yig‘ish, qayta ishlash va tahlil qilish.

Ma’lumotlardan oqilona foydalanish: Ma’lumotlar tahlili natijalarini qaror qabul qilish jarayonida to‘g‘ri qo‘llash.

Ma’lumotlarga asoslangan madaniyatga axloqni shakllantirishda ma’lumotlar mutaxassislari duch keladigan muammolar quyidagilarni o‘z ichiga oladi:

Ma’lumotlarning subyektiv tahlili: Tahlil natijalari mutaxassisning shaxsiy qarashlari va tajribasiga bog‘liq bo‘lishi mumkin.

Ma’lumotlarning noto‘g‘ri interpretatsiyasi: Tahlil natijalarini noto‘g‘ri tushunish va ulardan noto‘g‘ri xulosalar chiqarish.

Ma’lumotlardan manipulyatsiya qilish: Tahlil natijalarini muayyan maqsadlarga erishish uchun o‘zgartirish.

Axloqiy muammolarning yetarlicha e’tiborga olinmasligi: Ma’lumotlar tahlili bilan bog‘liq axloqiy muammolarga yetarlicha e’tibor bermaslik.

Ushbu muammolarni hal qilish uchun tashkilotlar quyidagi choralarni ko‘rishlari mumkin:

Axloq kodeksini ishlab chiqish: Ma’lumotlar bilan ishlashning axloqiy prinsiplarini belgilovchi hujjat.

Axloq bo‘yicha treninglar o‘tkazish: Xodimlarning ma’lumotlar tahlili bilan bog‘liq axloqiy muammolar haqida xabardorligini oshirish.

Mustaqil audit o‘tkazish: Ma’lumotlar tahlilining axloqiy prinsiplarga muvofiqligini tekshirish.

Shaffoflikni ta’minlash: Ma’lumotlar tahlili natijalarini barcha manfaatdor tomonlarga ochiq e’lon qilish.

Xulosa qilib aytganda, ma’lumotlarga asoslangan madaniyatni shakllantirishda ma’lumotlar mutaxassislari muhim rol o‘ynaydilar. Ularning mas’uliyati va duch keladigan muammolarini to‘g‘ri tushunish va hal qilish tashkilotning muvaffaqiyati uchun muhimdir.

Xulosa

Ushbu ishda avtomobilь sotuvi bilan shugʼullanuvchi kompaniyada katta maʼlumotlarni tahlil qilishning nazariy va amaliy jihatlari kompleks tarzda koʼrib chiqildi. Katta maʼlumotlarning asosiy tushunchalari, ularning 5V kabi muhim xususiyatlari (hajm, tezlik, xilma-xillik, ishonchlilik va qiymat), shuningdek, ular bilan ishlash bosqichlari, jumladan, maʼlumotlarni yigʼish, filtrlash, qayta ishlash, tozalash, vizualizatsiya qilish, tahlil qilish, soʼrovlar, foydalanish va saqlash kabi jarayonlar tahlil qilindi. Tashkilotda katta maʼlumotlarni boshqarishning muhim strategiyalari, masalan, maʼlumotlar koʼli va maʼlumotlar ombori kabi yondashuvlarning afzalliklari va kamchiliklari ham muhokama qilindi.

Shuningdek, maʼlumotlarga asoslangan qaror qabul qilishning ahamiyatiga alohida eʼtibor qaratildi. Bunday qarorlarning tashkilot uchun keltiradigan afzalliklari, jumladan, doimiy takomillashtirish va rejalashtirish, hamkorlikdagi qarorlar qabul qilish, xarajatlarni kamaytirish, real vaqtdagi tushunchalar, yangi imkoniyatlarni baholash, raqamli savodxonlikni oshirish va maʼlumotlarga asoslangan madaniyatni shakllantirish kabi jihatlar batafsil yoritildi. Qaror qabul qilish jarayonida yuzaga kelishi mumkin boʼlgan muammolar, masalan, nomuvofiq va nostandart maʼlumotlar, tavsiflovchi va bashorat qiluvchi tendentsiyalar oʼrtasidagi muvozanat kabi masalalar ham tahlil qilindi.

Maʼlumotlarni tayyorlash va manipulyatsiya qilishning muhimligi, shuningdek, manfaatdor tomonlarning ehtiyojlarini qondirish uchun statistik usullarni tanlashning ahamiyati alohida taʼkidlandi. Maʼlumotlarga asoslangan madaniyatga axloqni shakllantirishda maʼlumotlar mutaxassislarining roli, masʼuliyati va duch keladigan muammolari ham koʼrib chiqildi.

Ushbu tadqiqot natijalari avtomobilь sotuvi bilan shugʼullanuvchi kompaniyalarga katta maʼlumotlardan samarali foydalanish va operatsion samaradorlikni oshirish boʼyicha amaliy tavsiyalar beradi. Kompaniyalar maʼlumotlar sifatini taʼminlash, maʼlumotlar xavfsizligiga eʼtibor qaratish, malakali mutaxassislarni jalb qilish va axloqiy printsiplarga rioya qilish orqali katta maʼlumotlarning potentsialidan toʼliq foydalanishlari mumkin. Bu esa oʼz navbatida, biznes jarayonlarini optimallashtirish, mijozlar bilan munosabatlarni yaxshilash, sotuvlarni oshirish, raqobatda ustunlikka erishish va xavflarni samarali boshqarishga yordam beradi.

ADABIYOTLAR RO'YXATI

1. Frank, A., Knopp, B.M., Beaghley, S., Watson, M. and Orrie, R. (2016). *Defining the Roles, Responsibilities, and Functions for Data Science Within the Defense Intelligence Agency*. <https://www.researchgate.net/publication/310325102_Defining_the_Roles_Responsibilities_and_Functions_for_Data_Science_Within_the_Defense_Intelligence_Agency>
2. ‌ Columbus, L. (2014). Ten Ways Big Data Is Revolutionizing Manufacturing. *Forbes*. [online] 29 Nov. Available at: [https://www.forbes.com/sites/louiscolumbus/2014/11/28/ten-ways-big-data-is-revolutionizing-manufacturing/.](https://www.forbes.com/sites/louiscolumbus/2014/11/28/ten-ways-big-data-is-revolutionizing-manufacturing/)
3. ‌ McKinsey (2011). *McKinsey Global Institute*. [online] Available at: <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.pdf>.
4. ‌ DAMA International (2017) *DMBOK: Data Management Body of Knowledge*. 2nd ed. Bradley Beach, NJ: Technics Publications.
5. Databricks (2022) *The Data Lakehouse Architecture*. Available at: <https://www.databricks.com/> (Accessed: 20 January 2024).
6. Hyde, J. (2021) *Data Pipelines Pocket Reference*. Sebastopol, CA: O’Reilly Media.
7. Inmon, W. H. (2005) *Building the Data Warehouse*. 4th edn. Indianapolis, IN: Wiley.
8. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. and Becker, B. (2013) *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd edn. Indianapolis, IN: Wiley.
9. Loshin, D. (2009) *Master Data Management*. Amsterdam: Elsevier/Morgan Kaufmann.
10. Smith, J. and Jones, A. (2023). The Impact of Real-Time Data Analytics on Automotive Supply Chain Efficiency. *Journal of Automotive Technology*, 15(2), pp.123-145.
11. European Commission. (2020). *A European strategy for data*. [online] Available at: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066](https://www.google.com/url?sa=E&source=gmail&q=https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066) [Accessed 20 October 2023].
12. Marr, B. (2018). *How Big Data Is Transforming The Automotive Industry*. *Forbes*. [online] Available at: [invalid URL removed] [Accessed 20 October 2023].
13. Provost, F. and Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media