







ANALIZA DANYCH PLATFORMY E-COMMERCE OLIST WZBOGACONEJ O DANE BRAZYLIJSKICH MIAST

Aleksander Stepaniuk 272644

ETAP I: POZYSKANIE DANYCH

Źródła danych (Kaggle):

- Olist CSV (orders, items, payments, reviews, customers, sellers, products)
- Brazylijskie miasta CSV (populacja, powierzchnia, GDP)

▲ CITY 	▲ STATE 	# CAPITAL 	# IBGE_POP 	# AREA 	# GDP_CA 
Name of the City	Name of the State	1 if Capital of State	Resident Population Regular Urban Planning	City area (square kilometers)	Gross Dom per capita
5573 unique values	defesa [null] Other (2184) 35% 25% 39%	2465 total values	5573 total values	5573 total values	5 total
Abadia De Goiás	GO	0	5300	147.26	20664.57
Abadia Dos Dourados	MG	0	4154	881.06	25591.7
Abadiânia	GO	0	10656	1,045.13	15628.4
Abaeté	MG	0	18464	1,817.07	18250.42
Abaetetuba	PA	0	82956	1,610.65	8222.36
Abaíara	CE	0	4538	180.08	6370.41

ETAP 2: PROCES ETL

-
- The screenshot displays five SSIS Catalog job packages, each with a detailed flowchart of its tasks:
- SC_CreateStageTables:**
 - Task: **CreateStageSchema**
 - Parallel Path 1: **CreateStageOrders** → **CreateStageOrderItems** → **CreateStagePayments** → **CreateStageReviews** → **CreateStageCustomers** → **CreateStageSellers** → **CreateStageProducts**
 - Parallel Path 2: **CreateStageOrderClean** → **CreateStageCustomersClean** → **CreateStageSellersClean** → **CreateStageProductsClean**
 - SC_LoadStageData:**
 - Parallel Path 1: **DFT_LoadOrdersStage** → **DFT_LoadOrderItemsStage** → **DFT_LoadPaymentsStage** → **DFT_LoadReviewsStage** → **DFT_LoadCustomersStage**
 - Parallel Path 2: **DFT_LoadProductCategoryNameTranslationStage** → **DFT_LoadCleanStage** → **DFT_LoadProductsStage** → **DFT_LoadSellersStage**
 - SC_LoadFinalData:**
 - Parallel Path 1: **Load_MonthDim** → **Load_WeekdayDim** → **Upsert_FactOrders** → **Upsert_PaymentDim**
 - Parallel Path 2: **Upsert_TimeDim** → **Upsert_CustomerDim** → **Upsert_SellerDim** → **Upsert_ProductDim** → **Upsert_ReviewDim**
 - SC_CleanStage:**
 - Task: **DFT_OrdersClean** → **DFT_CustomersClean** → **DFT_SellersClean** → **DFT_ProductsClean**
 - SC_CreateFinalSchemaTables:**
 - Task: **Create Schema**
 - Parallel Path 1: **CreateMonthDim** → **CreateWeekdayDim** → **CreateTimeDim** → **CreateCustomerDim** → **CreateSellerDim** → **CreateProductDim**
 - Parallel Path 2: **CreateFactOrders** → **CreateReviewDim** → **CreatePaymentDim**
- A progress indicator on the right shows 60% completion.

ETAP 3: KOSTKA OLAP

Wymiary i fakty:

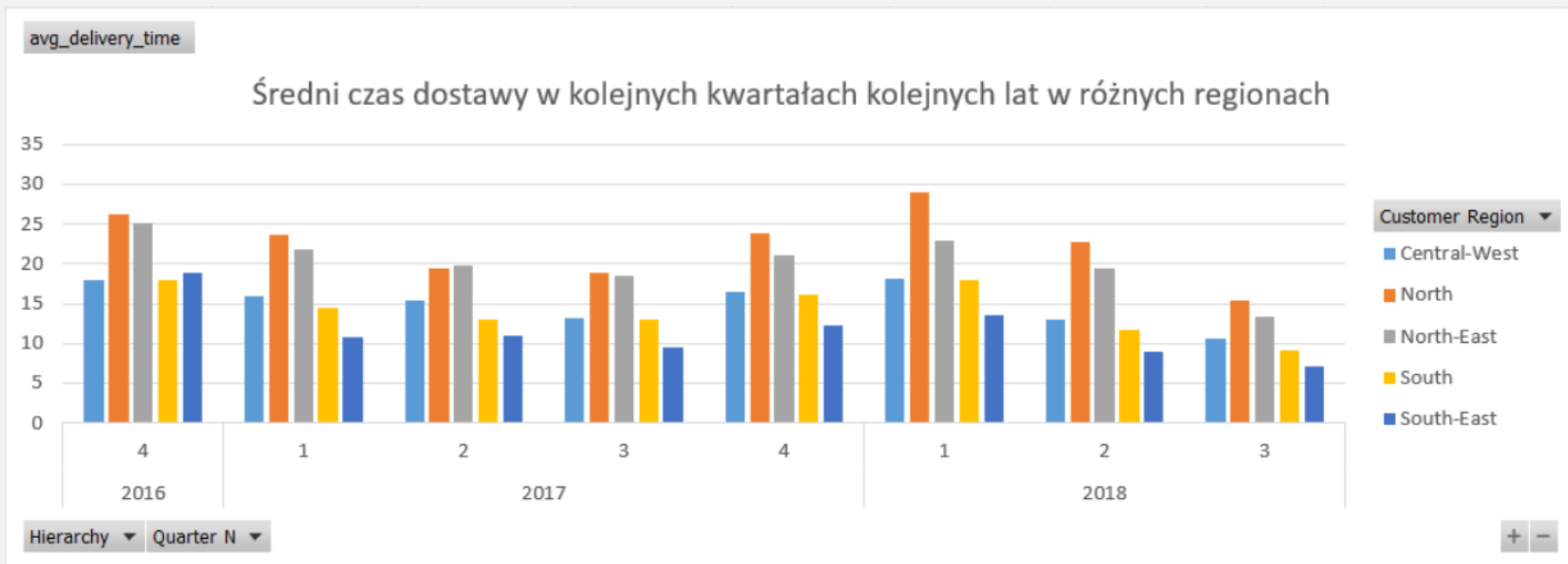
- TimeDim (rok, kwartał, miesiąc, dzień tygodnia, godzina, weekend)
- CustomerDim, SellerDim (miasto, stan, region, gęstość zaludnienia miasta, powierzchnia, gdp per capita, populacja)
- CustomerDim posiada także dodatkową miarę kalkulowaną „tag” mówiącą czy jest on nowym czy powracającym klientem.
- ProductDim (kategoria produktu przetłumaczona na język angielski), PaymentDim (typ płatności), ReviewDim
- FactOrders (każdorazowy zakup pojedynczego produktu)

DEMONSTRACJA WYNIKÓW

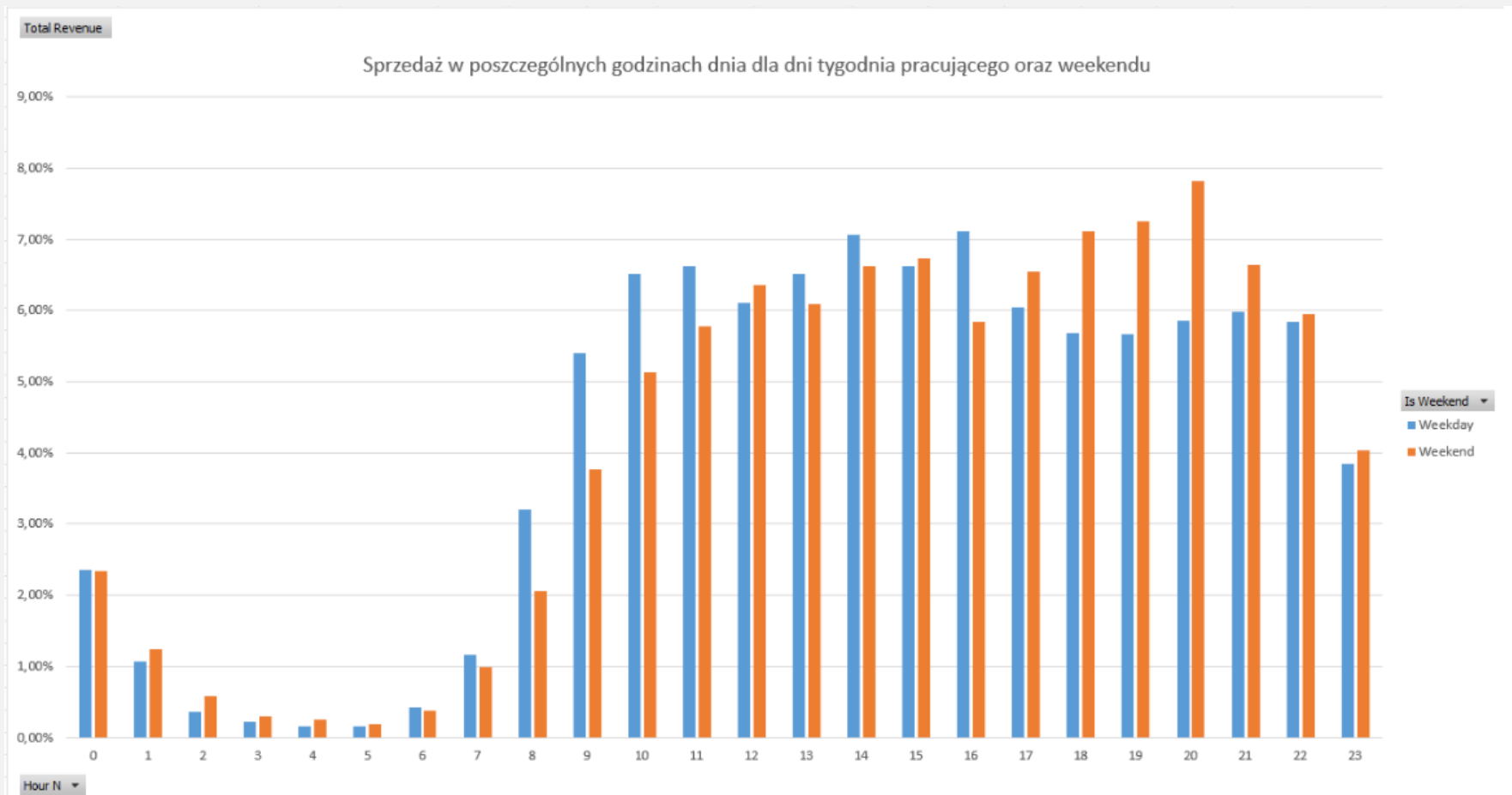
Wybrane analizy:

1. Średni czas dostawy w kolejnych kwartałach w różnych regionach
2. Sprzedaż w poszczególnych godzinach dnia dla dni tygodnia pracującego oraz weekendu
3. Średni czas dostawy dla miast o różnych powierzchniach w różnych miesiącach

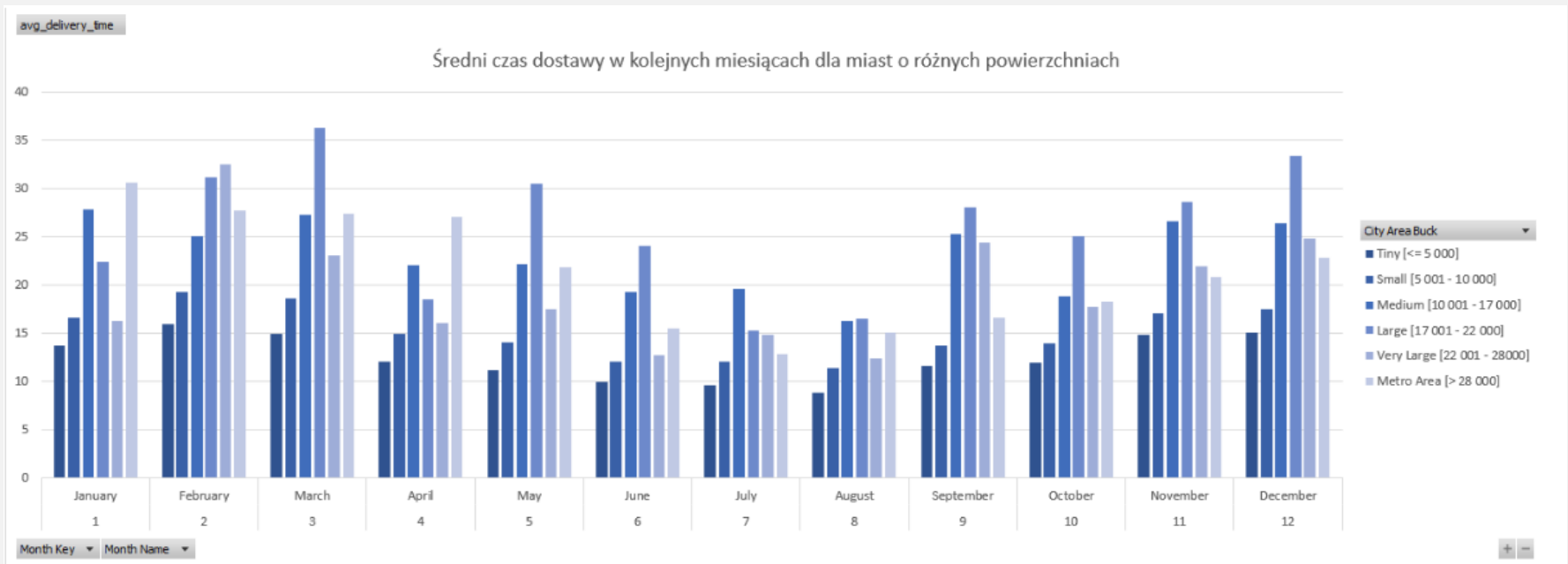
ŚREDNI CZAS DOSTAWY W KOLEJNYCH KWARTAŁACH W RÓŻNYCH REGIONACH



SPRZEDAŻ W POSZCZEGÓLNYCH GODZINACH DNIA DLA DNI TYGODNIA PRACUJĄCEGO ORAZ WEEKENDU



ŚREDNI CZAS DOSTAWY DLA MIAST O RÓŻNYCH POWIERZCHNIACH W RÓŻNYCH MIESIĄCACH



DZIĘKUJĘ ZA UWAGĘ

Aleksander Stepaniuk 272644