Lottery Sales in Texas

Business Analysis Template

Gabriela Cretu

June 2025

Contents

1 Business Description

1.1 Business Background

Lottery games, or luck-based games as we might call them, are part of a large and profitable industry around the world. People keep playing because of the strong desire to win big—sometimes just by scratching a ticket or picking a few lucky numbers. These games continue to be popular in many places.

In the United States, the lottery market is very large, and Texas stands out as one of the top states in terms of ticket sales. That's why this project focuses specifically on the Texas lottery. There are still many important questions to explore: Which games are the most or least profitable? What types of games do people in Texas prefer? Are instant-win scratch games more popular, or do players prefer waiting for draw results?

To find answers, it is important to understand how consumers behave, how well different games perform, and how retailers contribute to sales. This can be done by collecting and analyzing structured sales data from across Texas, which will help reveal useful insights and support better business decisions in this growing industry.

1.2 Problems Due to Poor Data Management

Poor data management can significantly hinder success in the lottery business. Without proper data, it's difficult to know which games are doing well and which ones are not. If you don't use tools that help you collect and analyze sales information, you won't be able to understand what players want or how retailers are performing.

In a competitive market like the Texas lottery, not having the right data can lead to missed opportunities, poor planning, and less profit. To stay competitive and make smart decisions, it's important to manage data properly and use it to create effective strategies.

1.3 Benefits from Implementing a Data Warehouse

Using a data warehouse can help solve the problems mentioned above. Implementing a data warehouse can answer important questions like:

- Which lottery games generate the most revenue?
- Which games have the widest range of sales across different retailers?
- Are there clear patterns in player preferences for instant-win versus draw-based games?

Further analysis of the data can also help to:

- Understand how sales vary by region or retailer type.
- Identify trends in customer behavior over time.
- Improve marketing strategies by targeting popular games.
- And many other useful insights.

By collecting and organizing data in one place, a data warehouse makes it easier to analyze sales, support better decision-making, and improve overall business performance.

2 Datasets Description

2.1 Scratch-Based Lottery

This dataset contains detailed transactional data for instant-win scratch tickets sold in Texas. Unlike the draw-based lottery, the assumption here is that a customer can purchase up to three tickets per transaction but can only win once per transaction because prizes are determined by ticket sellers. The prize value is represented by the `base_prize_scaled` attribute in the `DimPrizeRule` table.

Fact Table: ScratchTicketSalesFact

- **sale_id**: unique identifier for the sales record (PK)
- **date_surrogate_id**: foreign key to DimDate (FK)
- **game_surrogate_id**: foreign key to DimScratchGame (FK)
- **retailer_surrogate_id**: foreign key to DimRetailer (FK)
- **customer_surrogate_id**: foreign key to DimCustomer (FK)
- **employee_surrogate_id**: foreign key to DimEmployee (FK)
- **ticket_price**: price of a single scratch ticket
- **tickets_bought**: number of tickets purchased in the transaction (up to 3)
- **sale**: total transaction amount (ticket_price * tickets_bought)

Dimension Table: DimPrizeRule

- **prize_rule_surrogate_id** (PK)
- prize_rule_id
- game_id
- winning_type_id
- base_prize_scaled

Dimension Table: DimScratchGame

- **game_surrogate_id** (PK)
- scratch_game_number
- ticket_price
- average_odds
- average_odds_probs

Dimension Table: DimDate

- **date_surrogate_id** (PK)
- date_id
- fiscal_year
- fiscal_month
- fiscal_month_name_number

Dimension Table: DimRetailer

- **retailer_surrogate_id** (PK)
- retailer_license_number
- retailer_location_name
- retailer_number_and_location_name
- retailer_location_address_1
- retailer_location_city
- retailer_location_state
- retailer_location_zip_code
- retailer_location_county
- owning_entity_retailer_number
- owning_entity_retailer_name

Dimension Table: DimCustomer

- **customer_surrogate_id** (PK)
- customer_id
- customer_name
- customer_email
- customer_phone_number
- customer_street_address
- city
- state

- zip_code
- date_of_birth
- gender
- job_title
- company

Dimension Table: DimEmployee

- **employee_surrogate_id** (PK)
- employee_id
- employee_name
- email
- phone_number

2.2 Draw-Based Lottery

The dataset captures detailed information about draw-based lottery games. Unlike the previous scratch game, our assumption this time is that a player could win multiple times, as unlike the other one where his win is in the hands of the ticket seller, this time if he is sure of the right numbers(premonition), he could fill all three tickets and win big.

Fact Table: DrawTicketSalesFact

- **sale_id**: unique identifier for the sales record (PK)
- **date_surrogate_id**: foreign key to DimDate (FK)
- **game_surrogate_id**: foreign key to DimDrawGame (FK)
- **retailer_surrogate_id**: foreign key to DimRetailer (FK)
- **customer_surrogate_id**: foreign key to DimCustomer (FK)
- **employee_surrogate_id**: foreign key to DimEmployee (FK)
- **chosen_numbers_combination_surrogate_id**: foreign key to DimArrayCombinations (FK)
- **chosen_numbers_combination_surrogate_ids**: IDs of individual number surrogates

- **winning_numbers_combination_surrogate_id**: foreign key to DimCombinations (FK)
- **winning_combination_id**: original flat identifier for winning combination
- **ticket_price**: price of each ticket
- **ticket_sales**: total sale amount
- **tickets_bought**: number of tickets purchased in transaction
- **transaction_date**: date and time of the transaction
- **chosen_ids**: raw list of selected number IDs
- **winning_ticket_count**: number of winning tickets in transaction
- **is_winning_ticket**: boolean flag for win
- **winning_prize**: prize category or tier
- **amount_won**: total amount won

Dimension Table: DimDrawGame

- **game_surrogate_id** (PK)
- game_id
- game_category
- ticket_price

Dimension Table: DimCombinations

- **winning_numbers_combination_surrogate_id** (PK)
- winning_combination_id
- winning_numbers_combination
- game_category

Dimension Table: DimArrayCombinations

- **chosen_numbers_combination_surrogate_id** (PK)
- chosen_numbers_combination_surrogate_ids
- chosen_numbers_combination
- game_category

Dimension Table: DimDate

- **date_surrogate_id** (PK)
- date_id
- fiscal_year
- fiscal_month
- fiscal_month_name_number

Dimension Table: DimRetailer

- **retailer_surrogate_id** (PK)
- retailer_license_number
- retailer_name
- retailer_number_location_name
- retailer_address_1
- retailer_city
- retailer_state
- retailer_zip_code
- retailer_county
- owning_entity_number
- owning_entity_name
- owning_entity_chain_head_name

Dimension Table: DimCustomer

- **customer_surrogate_id** (PK): surrogate key, unique identifier for the customer
- **customer_id**: original customer ID from source system
- **customer_name**: full name of the customer
- **customer_date_of_birth**: birth date of the customer
- **customer_gender**: gender of the customer
- **customer_city**: city of residence
- **customer_state**: state of residence
- **customer_zip**: postal code / ZIP code

Dimension Table: DimEmployee

- **employee_surrogate_id** (PK)
- employee_id
- employee_name
- employee_email
- employee_phone_number

2.3 Proposed Use Cases

- Detect top-performing game types and retailers across counties.
- Build dashboards to compare scratch vs. draw-based sales performance.
- Track monthly growth, returns, and adjustments to enhance planning.
- Identify underperforming games and regions needing marketing support.