Fun with Data Analysis and BI

Real-world applications of analytical techniques

Nitin Sethi

www.bpbonline.com

First Edition 2025

ISBN: 978-93-55519-269

LIMITS OF LIABILITY AND DISCLAIMER OF WARRANTY

The information contained in this book is true to correct and the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but publisher cannot be held responsible for any loss or damage arising from any information in this book.

All trademarks referred to in the book are acknowledged as properties of their respective owners but BPB Publications cannot guarantee the accuracy of this information.

Dedicated to

My beloved parents:

Late Mr. R.K. Sethi

Late Mrs. Sunita Sethi

and

*My wife **Shail** and my twins **Aaradhyaa** and **Kartikey***

About the Author

Nitin Sethi is a data analytics leader, expert in machine learning, NLP, and business intelligence, and a mentor and educator.

With over 17 years of experience in data science, analytics, and strategy planning across various industries, he specializes in transforming complex data into actionable insights that drive strategic business decisions. His expertise spans a wide array of tools and technologies, including Python, SQL, Tableau, Alteryx, and advanced machine learning techniques such as deep learning, and supervised and unsupervised learning.

In his current role as a Data Analytics Lead at Rio Tinto, he leads a global team focused on reporting, analytics, automation, and tool development. He has successfully developed and deployed large-scale machine learning models using AWS services and Google Cloud, applying advanced experimentation techniques to validate business impact.

Previously, he spent a decade at McKinsey & Co., where he honed his skills in managing Data Analytics, conducting sentiment analysis using deep learning, and generating key performance indicators for diverse geographies and demographics.

Beyond his professional work, he has over a decade of experience teaching data analytics, machine learning, and artificial intelligence, with teaching roles at prestigious institutions like UT Austin and MIT. He is passionate about mentoring the next generation of data professionals and fostering a culture of data-driven decision-making.

About the Reviewers

- **Divit Gupta** is the Director and Product Advocate at Oracle Corp USA, dedicated to revolutionizing Cloud Computing, AI, and Data Science. Not only an accomplished IT professional, Divit is also a visionary leader shaping the future of technology. As a Senior Member of IEEE, he aims to empower over a million individuals worldwide with advanced skills in analytics, machine learning, and AI. With over two decades of experience, Divit is recognized as a prominent speaker and thought leader in emerging technologies.

  Divit engages audiences through his popular podcasts, including "Tech Talk with Divit," "Live Labs with Divit," and "Cloud Bites with Divit." He serves as a correspondent for Oracle TV's Cloud World and has presented on Oracle Database technology at Oracle Cloud World FY 2023. His expertise is also showcased in his books, "The Cloud Computing Journey," "The Potential of Generative AI," and "DBA Essentials 21C."

  Internationally, Divit is committed to knowledge-sharing, actively participating in conference talks, contributing to technical blogs, and writing Whitepapers. His insights are featured in leading newspapers and technology magazines worldwide. A highly skilled technologist, Divit holds professional certifications from Microsoft, Oracle, AWS, Databricks, and a post-graduate certificate from Harvard Business School Online.

  Divit's mission transcends personal achievement. He is dedicated to uplifting others and fostering their pursuit of excellence, with the overarching goal of creating a skilled and inclusive future in the tech industry.

- **Naresh Kumar Miryala**, a distinguished engineering leader at Meta, possesses an extensive background in cloud and platform engineering honed over nearly two decades in the field. His deep understanding of both technical and business intricacies empowers him to pioneer innovative solutions spanning diverse domains such as database systems, large-scale backend infrastructure, multi-cloud environments (AWS/GCP/OCI/Azure), automation, cloud infrastructure, DevOps, Kubernetes, and Elasticsearch.

  Naresh is an adept engineering executive and technologist with extensive experience in working with cloud migrations, infrastructure implementation, database management, ERP solutions, DevOps deployments.

Naresh is highly experienced in cloud migrations, particularly involving databases (Oracle/Exadata/MySQL/Postgres) and applications (EBS/Fusion/EPM/GTM). He played a pivotal role in ensuring their seamless execution. His experience in implementing infrastructure, databases, ERP solutions, and deployments has granted him a comprehensive understanding of the multifaceted technical and business challenges intrinsic to such projects.

Naresh's affiliations include membership in IEEE, OATUG, AIM leadership council, and fellowship at RSA. He holds certifications as a professional in Cloud and Database platforms, and actively engages as a blogger, tech reviewer, and frequent speaker in international conferences.

❖ **Natarajan R** has spent over 10 years in the IT industry, with experience in Unified Communications and Collaboration. He started as a Network Engineer for Trigent Software Limited and is currently working as a Unified Communications Engineer at Accenture Solutions Private Limited.

Natarajan R has authored books on Artificial Intelligence, Cloud Computing, and Blockchain Technology. He has also published several articles in Journals and is a reviewer for technical journals (IEEE). Natarajan R holds a C.Eng from the Institution of Engineers(India).

Acknowledgement

There are a few people I want to thank for the continued and ongoing support they have given me during the writing of this book: my in-laws, my brother, and my family for continuously encouraging me to write the book — I could have never completed this book without their support.

I offer my deepest gratitude to Sathya Sai Baba and Shirdi Sai Baba, whose divine presence and teachings have been a constant source of inspiration and guidance in my life. I am also profoundly thankful to Ruzbeh N. Bharucha for his mentorship and wisdom, which played a pivotal role in shaping this work. Their blessings, guidance, and support have been instrumental in helping me write this book, and I dedicate this work to them with heartfelt reverence.

My gratitude also goes to the team at BPB Publications for being supportive enough to give me quite a long time to finish the book.

Preface

This book will enable you to grasp SQL from a very basic level, covering installation, how to set an environment, the route to building the database, how to build the tables, and how you can set the queries on the tables end-to-end. This will give you hands-on experience with the SQL queries master database. This book will also teach you the fundamental theory, and from the theory, it will take you to the practical approach. You will have a full project where we have written the code for you, so if you just copy-paste, you will understand what is happening, and you will also be able to see the output at the end of the chapter. The book will also include some MCQ with solutions to help you in testing your knowledge of a particular concept.

Using the pragmatic approach, this project and ideas will assist you in revealing the full idea. Then, in one of the cases, we also talk about market analysis. We are trying to understand the stock market data and creating clusters on the same to look in detail and to understand the trend of the market, which will help you in creating such clusters that will be able to solve business problems and also if you want to improvise more on the model. You will be able to reach a good amount of accuracy, which can help individuals understand how stocks are performing. Also, for this one, we are using Python as the platform.

Then, we discuss the Churn analysis on the customer data set. We are trying to do this analysis to understand customer behavior and why they are leaving a particular product or a particular brand and moving to another one using a machine learning approach, and we will show how we can use Python as a platform.

Even if you are not using Python as a tool for visualization, tableau will be able to help you drag and drop your columns and create multiple visualizations without writing a single code. We have done hands-on to help you start the project from scratch and look at the result. We are talking about Tableau as a tool and how to install the public version. We will also provide the data set to create visualizations.

Chapter 1: E-Ticket Booking - This chapter is designed to give beginners a thorough understanding of developing an eTicket project using SQL. It covers all critical aspects, from creating a database schema to data restoration, aiming to equip readers with essential SQL programming skills needed for effective data management and manipulation in databases. The SQL-based eTicket project is particularly relevant for event management

companies, entertainment venues, and sports organizations, enabling them to efficiently manage events, sell tickets, and track sales. With this system, users can purchase tickets online, receive them via email or SMS, and scan them at the venue, reducing the need for paper tickets, saving time, and improving the overall user experience. SQL programming empowers these organizations to manage and manipulate data efficiently, ensuring the seamless operation of the eTicket system.

Chapter 2: Creating Games on Python - In this chapter, our primary objective is to introduce readers to three engaging digital game projects: Number Guessing, Rock Paper Scissors, and Dice Roll Generator. These projects serve as practical demonstrations of Python programming concepts, emphasizing interactivity, user input handling, and random number generation. By actively participating in the creation of these games, readers will gain valuable hands-on experience in coding logic, user interfaces, and game mechanics.

Chapter 3: Introduction to Sentiment Analysis - The chapter provides a comprehensive introduction to sentiment analysis, covering its definition, uses, benefits, and the tools and libraries commonly employed in the field. We will explore the underlying concepts, techniques, and methodologies used in sentiment analysis, and discuss real-world use cases to illustrate its practical applications.

Chapter 4: Sentiment Analysis on E-Commerce: Product Reviews - This chapter aims to equip readers with the skills to conduct sentiment analysis on e-commerce product reviews. By its conclusion, readers will understand the significance of sentiment analysis in e-commerce, learn to acquire relevant datasets, and handle missing data using appropriate methods. Mastering these techniques will enable readers to analyze and interpret customer sentiments, helping businesses improve products, enhance customer experiences, and make strategic decisions. By identifying trends and monitoring brand reputation, companies can optimize product development, marketing, and customer experience strategies, ultimately increasing customer satisfaction and loyalty and gaining a competitive edge in the e-commerce market.

Chapter 5: Sentiment Analysis on X - X sentiment analysis provides businesses with a valuable tool to proactively manage their brand reputation. By monitoring and analyzing sentiments expressed on X, companies can gain real-time insights into customer perceptions of their brand. Positive sentiments indicate customer satisfaction and loyalty, while negative sentiments highlight areas for improvement. This analysis helps businesses identify issues, address concerns, and make data-driven adjustments to their strategies. Additionally, comparing brand sentiment with competitors allows companies to understand their market positioning, which allows for crisis management by promptly

addressing negative sentiments. Ultimately, X sentiment analysis enables businesses to strengthen customer relationships, enhance products and services, and maintain a positive brand image in the digital landscape.

Chapter 6: Stroke Prediction - The objective of this chapter is to demonstrate how machine learning techniques can be employed in stroke prediction. By analyzing a dataset containing various health-related features, we aim to develop accurate models that can identify individuals at risk of experiencing a stroke. The models trained in this chapter, such as Random Forest, Gradient Boosting, and Logistic Regression, will help healthcare professionals make informed decisions and provide targeted interventions to high-risk patients. Through this use case, we aim to showcase the significance of data-driven approaches in healthcare, specifically in early detection and prevention of stroke, leading to improved patient outcomes and better allocation of medical resources.

Chapter 7: Movie Review Sentiment Analysis - Movie review sentiment analysis is a powerful tool for the film industry, providing valuable insights into audience perceptions. By analyzing sentiments expressed in reviews, filmmakers and studios can gain a deep understanding of audience reactions toward their movies. Positive sentiments indicate high levels of customer satisfaction and loyalty, allowing filmmakers to identify successful elements and capitalize on them in future productions. On the other hand, negative sentiments highlight areas for improvement, guiding filmmakers to address issues and enhance their storytelling techniques. Additionally, sentiment analysis helps studios assess the overall reception of their films and adjust marketing strategies accordingly. By leveraging sentiment analysis, the film industry can make data-driven decisions, tailor content to audience preferences, and ultimately create more engaging and successful movies that resonate with viewers.

Chapter 8: Stock Market Data Analysis - In the dynamic financial markets, this chapter equips investors with essential tools to leverage data-driven insights and advanced techniques. By exploring Exploratory Data Analysis (EDA) and clustering, readers can interpret complex financial data, identify stock groups with shared traits, and optimize portfolios based on risk tolerance. The chapter also emphasizes diversification to mitigate risks and enhance returns. Whether you're new to investing or experienced, this chapter empowers you to make more informed, confident, and successful investment decisions through data-driven strategies.

Chapter 9: Customer Data Analysis - This chapter is designed to provide readers with the skills and knowledge needed to analyze customer churn in the banking sector effectively. By exploring customer behavior and preferences through data analysis, readers will learn

to master EDA, understand the importance of customer segmentation using clustering for targeted marketing, and implement strategies for risk mitigation. The chapter also covers practical skills in data preprocessing, model training, evaluation, and cross-validation. By the end, readers will be equipped to make informed, data-driven decisions that enhance customer engagement and drive growth in a competitive business environment.

Chapter 10: Analyzing Sports Data in Tableau - In this chapter, we analyze historical IPL match data to identify patterns and trends in team performance. Our goal is to demonstrate how these insights can enhance the fan experience by highlighting standout moments and performances, allowing teams to create more engaging content. This personalized approach strengthens the bond between teams and their supporters. By utilizing data-driven analysis through Tableau, we show how this can transform decision-making in cricket, from optimizing team strategies and player selections to engaging fans more personally, ultimately giving teams a competitive advantage in the highly popular sport of cricket.

Chapter 11: Office Supplies Dashboard Using Tableau - By the end of this chapter, the readers will learn to navigate the realm of office supplies analytics with Tableau as the guiding tool, enabling readers to delve into the intricacies of data visualization. By mastering the creation of visualizations based on various columns and segments, understanding the nuances of calculations, and harnessing the power of filters, readers will gain the ability to unearth valuable insights from their data. Additionally, delve into the realm of action filters and automation to enhance dashboard interactivity and efficiency. By the chapter's end, readers will possess the knowledge and skills to construct engaging and informative office supplies dashboards, empowering them to optimize procurement strategies, identify trends, and make data-driven decisions.

Chapter 12: COVID Dashboard Using Tableau - To leverage data analytics and visualization tools, such as Tableau, to provide a comprehensive understanding of the COVID-19 pandemic's impact on a global scale. Through insightful analysis and interactive visualizations, the chapter aims to achieve the following objectives: Understanding pandemic trends, assessing healthcare systems, highlighting regional variances, exploring vaccination efforts, addressing socio-economic impacts, promoting data-driven decision-making, and fostering resilience and innovation.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/erv88rq

The code bundle for the book is also hosted on GitHub at
https://github.com/bpbpublications/Fun-with-Data-Analysis-and-BI.
In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1
E-Ticket Booking

Introduction

Structured Query Language (**SQL**) is a programming language for managing relational databases. It allows you to perform various tasks, such as creating and modifying database structures, inserting and retrieving data, and updating and deleting data. SQL is a standard language mostly used by database management systems, including MySQL, Oracle, and SQL Server.

SQL is an essential skill for data management professionals, including database administrators, data analysts, and developers. It allows you to efficiently manage and manipulate large amounts of data, providing valuable insights into business operations and decision-making.

The world is increasingly digital, and the eTicket system is one example of how technology is transforming traditional practices. Electronic ticketing, or eTicketing, is a system that uses electronic tickets instead of paper tickets. It allows users to purchase tickets online, receive them via email or SMS, and scan them at the venue. SQL is a programming language that is commonly used for managing and manipulating data in databases.

This book chapter will provide an overview of how beginners can develop an eTicket project based on SQL. It will cover various topics such as database schema creation, table designing and creation, relationship definition, data insertion, basic SQL queries writing,

data updating and deleting, views creation, queries optimization, and backup and data restoration.

Structure

In this chapter, we will learn the following topics:

- Environmental setup
- Relational database
- Tables, relationships, and primary keys
- Basic SQL syntax
- Project

Objectives

This chapter aims to provide beginners with a comprehensive understanding of how to develop an eTicket project based on SQL. It aims to cover all the essential aspects of creating an eTicket system, from database schema creation to data restoration. Additionally, it aims to help beginners develop skills in SQL programming, which will be essential for managing and manipulating data in databases. The eTicket project based on SQL is relevant for event management companies, entertainment venues, and sports organizations. These organizations can use the eTicket system to manage events, sell tickets, and track ticket sales. With the eTicket system, users can purchase tickets online, receive them via email or SMS, and scan them at the venue. This reduces the need for paper tickets, saves time, and enhances user experience. Using SQL programming, event management companies, entertainment venues, and sports organizations can manage and manipulate data effectively, ensuring the eTicket system runs smoothly.

Environmental setup

Before diving into SQL, you will need to set up a development environment. You will need to download and install a SQL server, a SQL client, and a sample database. You can download SQL Server Express and **SQL Server Management Studio (SSMS)** for free from Microsoft's website. You can also download a sample database, such as AdventureWorks or Northwind from Microsoft's website. Once you have installed the necessary software, you can begin practicing SQL by running queries and experimenting with different SQL statements.

Follow the steps mentioned below to download SSMS:

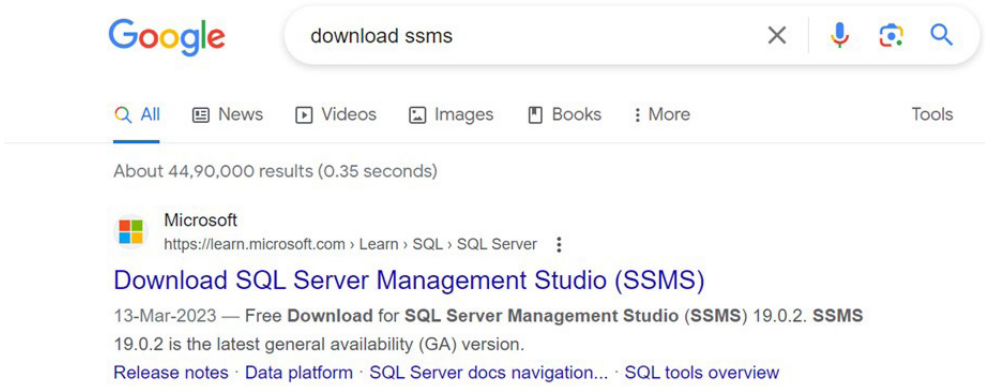

Figure 1.1: *The screen displaying the first link to click on is shown*

In the following figure, **Download SQL** Server is depicted:

Figure 1.2: *The instruction to click on the "Download SQL Server" option is depicted*

In the following figure, various services are displayed, providing an overview of the available options:

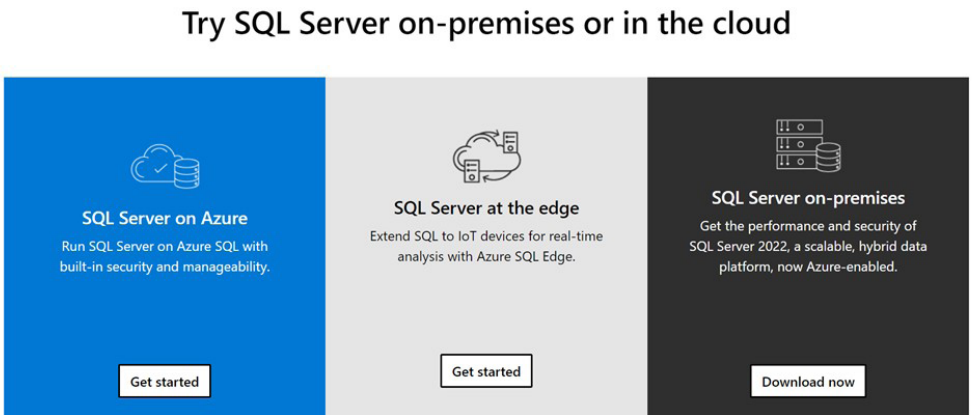

Figure 1.3: *The various services are displayed, providing an overview of the available options.*

In the next figure, required details can be filled in to proceed with the necessary steps:

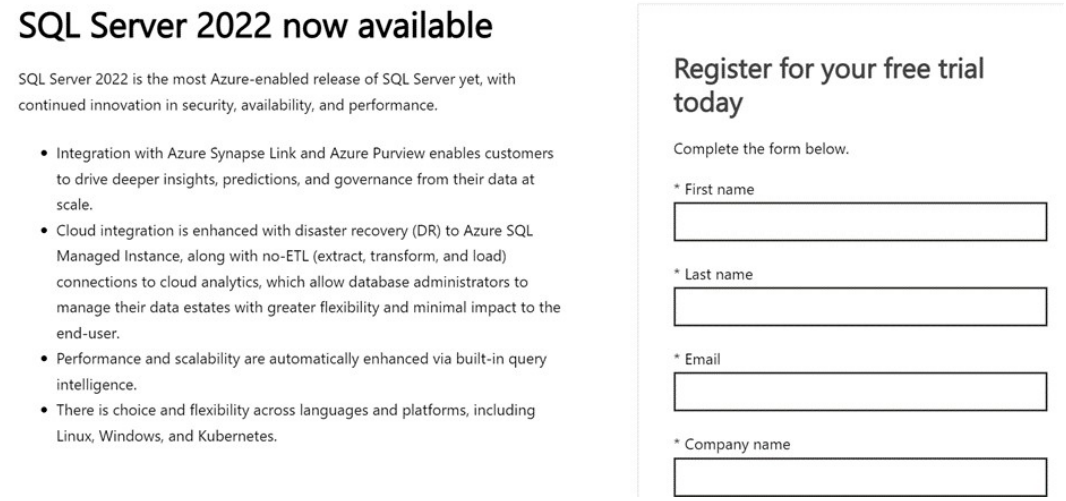

Figure 1.4: *The required details can now be filled in to proceed with the necessary steps.*

In the following figure, next step after filling in the required details, where you have to click on the **Download now** button is depicted:

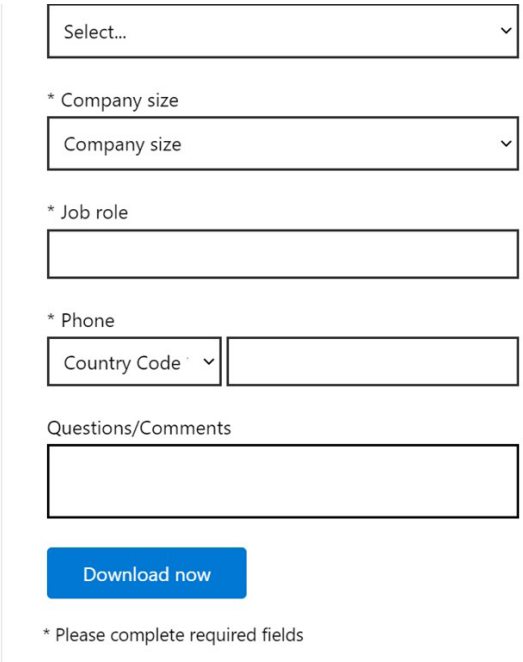

Figure 1.5*: The next step after filling in the required details, where you have to click on the "Download now" button*

The next step would be selecting **64-bit edition**:

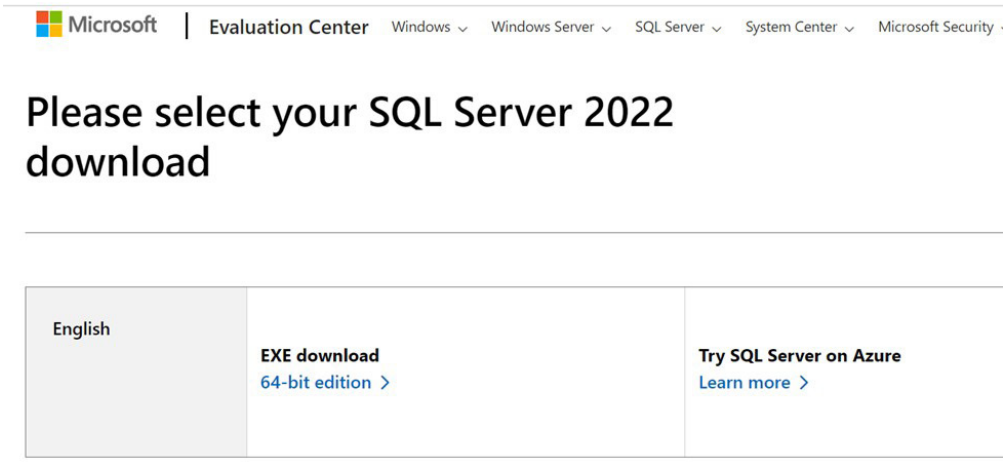

***Figure 1.6**: Perform the subsequent step, which involves selecting the 64-bit edition.*

In the next step, you will find the downloaded file in the **`Downloads`** folder:

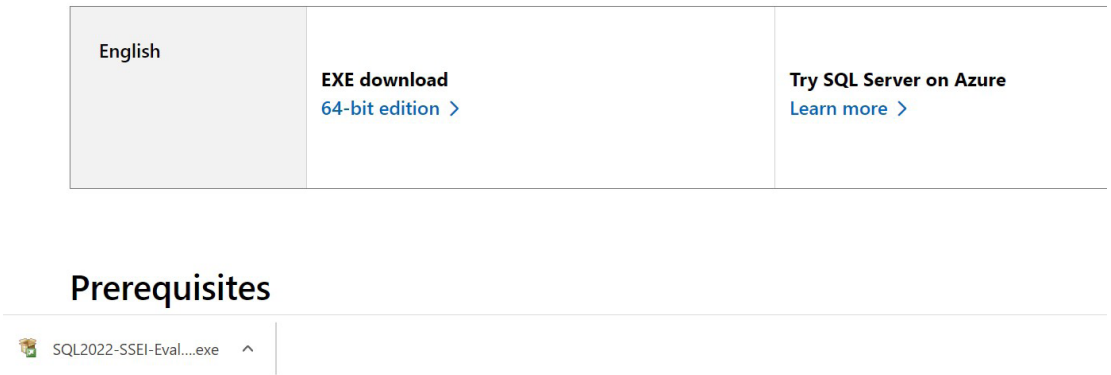

***Figure 1.7:** In this step, you will find the downloaded file in the downloads folder.*

Relational database

A relational database is a collection of data that is organized into tables. Each table contains rows (also called records or tuples) and columns (also called **fields**). The columns define the type of data that is stored in each row. The rows contain specific instances of that data.

Relational databases are used to store and manage data for a wide range of applications, including inventory management, customer relationship management, and financial transactions.