

Jun 4, 2025

Automated Cash Forecasting: Practical Insights from CRISP-DM

Streamlining analytics projects using CRISP-DM's six-phase approach.

Content Type: Article | Reading Time: 15 Minutes

In an era of increasing market volatility and the need for rapid adaptation, data analytics presents game-changing opportunities across industries, significantly boosting productivity and decision-making. Yet, finance and management accounting have been slower to embrace advanced analytics compared to fields like marketing or supply chain management. While the potential for transformation is clear, implementation remains complex, often due to a lack of practical guidance on how to approach such projects.

The Cross-Industry Standard Process for Data Mining (
CRISP-DM

) has become a leading framework for structuring analytics projects. Its adaptability and proven success across industries make it a valuable tool for finance professionals looking to integrate data-driven decision-making. However, while many studies outline

theoretical applications of CRISP-DM

, this article takes a different approach—showcasing how it plays out in the real world.

Through an in-depth examination of an automotive retailer's finance function, this article explores the practical application of CRISP-DM in cash forecasting. By walking through each phase of the process, from data understanding to deployment, it highlights key insights, unexpected challenges, and practical lessons that often go unnoticed in abstract frameworks. The goal is to provide finance professionals with a relatable, hands-on perspective on how analytics can enhance forecasting accuracy and efficiency, bridging the gap between theory and implementation.

Analytics framework

The CRISP-DM framework outlines the life cycle of a business analytics project through six phases:

- Business understanding, where the project's goals and requirements are defined.
- Data Understanding, involving initial data collection and familiarization.
- Data preparation, where data is cleaned and formatted for analysis.
- Modeling, where various algorithms are applied and tuned.
- Evaluation, assessing the model's performance against project objectives.
- Deployment, integrating the model into business processes.

These phases are arranged in a circular process model illustrated in Figure 1, representing the cyclical nature of applied data science. This iterative process often involves revisiting and refining steps as insights evolve and conditions change.

Liquidity planning and cash forecasting

At a leading European automotive retail group, founded in the mid-1940s and still family-run, data analytics is playing an increasing role in financial decision-making. The company, with annual revenue exceeding \$4 billion and a 30% market share in its home market, has expanded far beyond car import and retail business into leasing, repair, and fleet management.

Liquidity planning is crucial, as the car import business generates tens of millions of dollars in weekly cash flows. It is important to note that contrary to the process in the U.S., where cars are stock vehicles, in Europe, most cars are individualized orders from customers. After producing a car, the manufacturer issues an invoice to the organization as the car importer (cash outflow). After the delivery and sale of the car, the organization issues an invoice to the buyer (cash inflow). The timing of these cash flows is important, as global car manufacturing networks frequently have delays in delivery or uncertain demand for inventory cars, introducing a high level of fluctuation. With the outbreak of the COVID-19 pandemic, the car retailer started to track liquidity more closely, since having an overview and forecasting the near future cash flows is imperative for steering in times of extreme uncertainty.

For years, the company relied on manual Excel models to forecast weekly cash flows, using historical averages of delivery times, storage duration, and car prices. This process involved multiple departments (for example, group finance, import, and sales steering) and typically took five business days to complete.

The company recently adopted a company-wide data science platform to streamline collaboration and deploy large-scale business analytics projects. Recognizing the potential of its data and the need for a more accurate, automated cash forecasting solution, the company launched an initiative to leverage advanced analytics for improved liquidity planning.

Results: applying CRISP-DM to internal cash forecasting

As one of the company's first projects of this kind, there were no existing processes to build on. Therefore, the CRISP-DM model was chosen as a suitable framework for developing such a cash forecasting model from scratch.

Business understanding

Precisely formulated objectives and a thorough plan are essential for a project's success. Since the project team consisted of members from various internal departments, including group finance, individual business unit finance, IT, and external research team members, aligning on a shared understanding of business processes was a key priority. This alignment was primarily achieved through interviews with employees, focusing on the cash-triggering processes from customer car order to cash flow (order-to-pay). Conversations with teams from sales steering, logistics, and accounting provided crucial insights into supply chain dynamics, delivery and customs challenges, and variations in payment terms (for example, different manufacturing plants operating under different payment schedules, influencing cash flow timing).

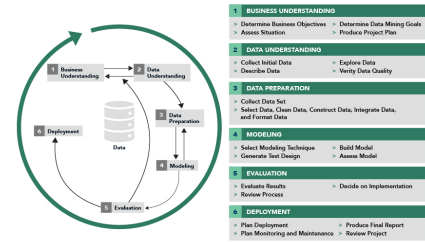


Figure 1: Phases and core components of the CRISP-DM process model.

A particularly revealing moment occurred during the development of the Driver Tree, which provided a structured representation of the key cash flow drivers in the retailer's import business. By mapping these drivers explicitly, team members gained a clearer understanding of the factors influencing cash flow timing. This representation became a valuable tool for aligning different perspectives across departments.

As part of this effort, a process mapping workshop helped clarify certain aspects of the cash flow process. Visualizing the end-to-end workflow made some interdependencies suddenly clear to the team. For example, controllers gained a new understanding of sales processes, while sales teams, in turn, saw the financial workflow in a new light. This explicit representation helped bridge gaps in perception, leading to more effective collaboration and decision-making.

To ensure continuity and knowledge retention, interviews were recorded and transcribed for future reference and to facilitate onboarding of new team members. The Driver Tree played a central role in consolidating these insights, serving as a reference point for key stakeholders. Based on these findings, three primary objectives were defined: time gain, to produce forecasts faster; efficiency gain, to make the process less resource-intensive; and quality gain, to improve forecasting accuracy.

Data understanding

The CRISP-DM framework postulates an iterative relationship between the data and business understanding phases. Given that the data originated from multiple organizational units, cross-checks for data compatibility were crucial. One particularly revealing challenge emerged regarding the unique vehicle identifier, which, as it turned out, wasn't so unique across the company. Different departments used different formats: one relied on a purely numeric code, while another used a mix of letters and numbers. Initially, discussions assumed everyone was referring to the same identifier, leading to confusion when data didn't align. Only through extensive cross-checks and follow-up discussions it was possible to clarify which identifier was used where and establish a unified approach for the analysis.

A primary component of the data understanding phase was exploratory data analysis. This involved checking for missing values, analyzing relationships with the target variable, and assessing the impact of major events like COVID-19 or the microchip supply crisis. For example, an analysis of the relevance of car brands in predicting cash flows revealed that brand significantly influences production and delivery times, ultimately affecting cash flow timing.

By better understanding the behavior of key value drivers, identified in the business understanding phase, the later steps of data preparation and modeling were significantly simplified, ultimately leading to better modeling outcomes. The process reinforced how crucial it was to resolve fundamental data interpretation issues early, preventing misalignment that could have compromised the entire analysis.

Data preparation

While the earlier phases focused on understanding the project and data, the data preparation phase was much more hands-on, transforming raw datasets into something usable for business analytics. This process required merging datasets from multiple departments, including production, vehicle orders, and customer data. However, not all data was relevant; for instance,

vehicles involved in intercompany transactions were excluded since they had no direct cash effect.

Feature engineering played a crucial role in shaping the data for analysis. Since the objective was to determine cash flow timing (the days between a vehicle order and its corresponding cash flow), date features were first standardized to ensure consistency across datasets and then transformed into time-span features by calculating the differences between relevant timestamps. One exemplary discovery during this phase was encountering vehicle production years as early as 1924 in the dataset, an obvious data error, but a reminder of why thorough data cleansing is crucial. Amusingly, this vehicle was still marked as under construction and had never been delivered, raising the question of whether it was the world's longest-running production delay.

Beyond basic corrections, broader market disruptions also had to be considered. Given the impact of COVID-19 and the microchip supply crisis, moving averages (up to 12 months) were applied to capture short-term fluctuations while distinguishing them from long-term trends. Additionally, simple regression and classification models imputed missing values, such as predicting production plants based on vehicle brand, type, and dates. This process exemplified the iterative back-and-forth, where refining and revisiting decisions were key, much like the CRISP-DM framework suggests.

Modeling

The modeling phase is highly iterative, often requiring hundreds of model and feature selection iterations before arriving at the final prediction model. Initially, certain features were expected to be strong predictors but later proved to have little impact, while other, less obvious variables turned out to be crucial. For instance, vehicle price, expected to have a significant impact on the dollar amount of cash inflows, had almost no effect. In contrast, a seemingly minor detail, the number of production updates in the prior weeks, turned out to be highly indicative of the upcoming cash flows. Frequent updates often signaled disruptions or logistical delays, influencing payment timelines more than the price itself. This iterative journey reinforced the importance of empirical testing over assumptions, some features that initially seemed crucial had to be discarded, while unexpected patterns emerged as key drivers. Success came not from relying on intuition alone, but from systematically testing, discarding, and refining until the optimal combination of models and features was found.

Real-world use cases like cash forecasting typically require multiple machine learning models. More than four distinct models were needed to handle the complexities of cash outflows, inflows, delivery dates, and vehicle prices. Additionally, time series models were employed to provide long-term cash flow projections. Various algorithms were trained using 10-fold cross-validation, with tree-based machine learning models delivering the best performance and lowest prediction error.

Evaluation

Before integrating a new model into business processes, ensuring it meets the initial business objectives is essential. An objective evaluation report is prepared for the project steering committee to decide on the model deployment or further iterations through the CRISP-DM. Reporting on model accuracy is crucial, but it should be contextualized against existing forecasting methods (like naïve or time series forecasts) to highlight the value added by the new analytics model. This involved comparing the machine learning-based forecasts to the previously manually planned cash flow forecasts. The new model excelled in all three objectives: reducing forecast time

from five days to one, increasing efficiency by automating forecasts, requiring validation by only one employee, and improving accuracy.

However, even though the model clearly outperformed manual forecasting, getting buy-in from business stakeholders proved to be an unexpected challenge. Many finance and planning teams were initially skeptical, relying on a machine instead of expert intuition felt like giving up control over a process that had always been judgment driven. The turning point came when historized datasets were used for back testing, demonstrating that the machine learning model would have flagged major liquidity risks weeks earlier than manual planning had. This tangible evidence of the model's advantage helped convince decision-makers, reinforcing the importance of not just building an accurate model, but also proving its real-world impact in a way that resonates with stakeholders.

To further build trust in the model, the car retailer tackles performance assessment through a rolling forecast evaluation technique to estimate future model performance. Rolling forecast evaluation simulates real-world forecasting by testing model performance over a specific horizon (for example, eight weeks) and time period (for example, one year). The previous test data is included in training the new model for each forecast, ensuring all available data is used. Future model performance is then estimated by averaging all test set error metrics.

Deployment

The deployment phase marks the final stage of CRISP-DM, focusing on determining a strategy for implementing the business analytics results. This involved setting up an automated pipeline on the data science platform. This pipeline enables the end user to upload weekly raw datasets and run the forecast, which takes about 1.5 hours. Afterward, the management accountant can access and analyze the results.

Initially, there was hesitation among management accountants to fully rely on the system-generated forecasts. Despite the model's proven accuracy, users cross-checked every result manually, sometimes even recreating parts of the forecast in Excel to verify its reliability. To address this, multiple options for consuming the forecast were provided. Outputs were made available as Excel exports for further analysis and aggregation with other business units, ensuring a familiar format for those still building confidence in the new system.

Additionally, an interactive dashboard was developed to provide human validation and actionable insights. This allowed users to filter and drill down on specific aspects, such as car brand-level forecasts, making it easier to verify individual components of the predictions. Over time, as accountants saw the forecasts consistently align with their expectations, trust in the automated process grew. The dashboard ultimately became a key tool in bridging the gap between automation and human decision-making, ensuring that insights were not just accurate but also actionable and trusted.

Conclusion and recommendations

The new forecasting approach significantly improves time, efficiency, and quality compared to the previous manual process, streamlining the cash forecasting process for the finance department. Beyond this, the entire organization gained valuable insights from this project, acquiring knowledge in implementing business analytics and laying the groundwork for other core processes

such as storage planning or sales staffing.

Using CRISP-DM effectively structures and streamlines the implementation of analytics applications, promoting continuous improvement. Key learnings from this application include the importance of clear objectives in the business understanding phase to then have decisive success criteria in the evaluation phase. Understanding stakeholders and project aspirations is crucial for this. A data dictionary proved invaluable for documentation, onboarding, and model maintenance in the data understanding phase. In the evaluation phase, splitting data into training, validation, and test sets, as well as testing different models, proved beneficial for assessing model performance. Additionally, historizing data early allowed for realistic back testing, ensuring the model's accuracy under near-real-world conditions. Since CRISP-DM is iterative, it's essential to remember that even successful models need continuous updates as business drivers evolve during deployment.

Summary of recommendations	
Business understanding	• Learn from existing forecasting practices as they usually provide valuable insights into key forecast drivers and common anomalies such as year-end effects. • Don't neglect external factors and data sources, including integrated supply chain data received from suppliers.
Data understanding	• Take the time to create a data dictionary – it will pay off in the long run. For example, it will help get new project team members up to speed quickly, or there might be a substantial time slack between the data understanding and evaluation phases where valuable information might be lost if a new CRISP-DM iteration is required. • Don't just drop or neglect missing data points; impute them with logical reasoning or additional models. Another strategy is to turn missing data points into a favor (a missing value is often also a predictive indicator – see “missing value indicators”).
Data preparation	• Split your dataset into training, validation, and test sets, and keep the test set strictly for the evaluation phase (only unseen data ensures an objective evaluation of your model). • Don't prepare your training, validation, and test sets separately – this saves time and ensures that you have all features available at the time of forecasting. For example, if features such as exchange rate forecasts are relevant for your model, you need to ensure that you have them available for the forecast horizon.
Modeling	• Combine a variety of models and modeling approaches to represent the complexity of real-world forecasting problems. • Consider the trade-off between the higher complexity of your modelling approach versus forecast run-time and long-term maintenance cost.
Evaluation	• Start early with the historization of datasets for back testing under near-real-world conditions. This is especially relevant if you work with continuously changing transactional datasets. • Don't stay vague in formulating and evaluating the objectives (success criteria) of your project (for example, “>X% accuracy improvement compared to the current method, measured as mean absolute error (MAE) over the time horizon Y).

Summary of recommendations	
Deployment	• Reengineering of existing business processes is an essential part of the deployment phase and requires strong stakeholder engagement. • Don't neglect recurring model maintenance, as model performance might deteriorate over time (for example, due to new business drivers).

Table 1: Summary of recommendations for applying CRISP-DM.

While these recommendations broadly apply to business analytics in finance and management accounting, implementation challenges may vary based on an organization's analytics maturity. CRISP-DM is best suited for well-defined analytics use cases with an overarching strategy and requires familiarity with iterative and incremental project management approaches (for example, Agile methods such as Scrum Framework (or Scrum)). Moreover, access to business analytics tools and technology for fast extraction, transformation, and loading (ETL) of data, experimentation with machine learning models, and scalable deployment capabilities is a prerequisite. With the rise of automated machine learning, the modeling phase will likely accelerate, reducing the need for deep data science expertise, particularly in finance departments, where less complex algorithms often suffice. This development will further strengthen the role of the management accountant in driving the implementation of business analytics.

About the authors

Dr. Alessandra Huydts is a Research Associate at the University of St. Gallen (HSG).

Klaus Moeller is a professor and chair of Controlling/Performance Management and director of the Hilti Lab for Integrated Performance Management at the University of St. Gallen. He's also a member of IMA's Switzerland Chapter.

Michael J. Weiser is a research associate and Ph.D. candidate at the University of St. Gallen and a member of IMA's Switzerland Chapter.

Want to learn more about IMA?

Our valuable courses and certifications help you stay ahead with the latest trends, research, and best practices in management accounting.

*Enter your email address**

Get more info

☐ I agree to the [IMA Privacy Policy](#) and would like to receive marketing emails.*

 [Email us at ima@imanet.org](mailto:ima@imanet.org)

 [Call us at \(800\) 638-4427](tel:(800)638-4427)

Company

- [About IMA](#)
- [IMA Leadership](#)
- [Our Certifications](#)
- [Press Releases](#)
- [Ethics Center](#)
- [Careers at IMA](#)

Support

- [Donate](#)
- [Knowledge Center](#)
- [Chapter Leader Portal](#)
- [Volunteering with IMA](#)
- [Sponsorships](#)
- [Contact Us](#)



[Privacy Policy](#) [Terms & Conditions](#) [Cookie Policy](#)

© 1997-2026 Institute of Management Accountants. All Rights Reserved.