日本SASユーザー会
フォーラム
2020

初心者からのSAS Viya

〇森岡　裕
(イーピーエス株式会社　統計解析部)

SAS Viya from a beginner's view.

Yutaka Morioka
Statistics Analysis Department 1, EPS Corporation

日本SASユーザー会
フォーラム
2020

要旨:

SAS Viyaの体験版の体験談.

生粋のコード書きである, ープログラマからみたSAS Viyaの所感

キーワード:SAS Viya

Disclaimer（おことわり）

SAS Viyaについては,既に多くのPaper・発表が公開されています.

SAS Viyaによって新しい知見を切り開いたり，実務に応用した例を示すものでなく

本発表はあくまで，1ユーザーとしての体験談の共有となります.

このセッションは息抜きです！！気軽にお聴きください

今年はSASユーザー総会がないのか…，テンション下がるなぁ…↘↘

SASユーザーフォーラムで発表の話を頂く↗↗

しかし，テーマとしてSAS Viya…

触ったことない…

深層学習・AIの実務経験もないような生粋のプログラマがはたしてViyaの発表をしてよいものだろうか…，悩む…↘↘↘↘↘↘↘↘↘↘

しかし･･･，大半のユーザーはviyaを使ったことがないはず．

機械学習・深層学習・AI についても，なじみのある方や経験・知見をお持ちの方ばかりではないはず･･･

すでにスペシャリストである方の発表は既に多くなされているが，案外，自分のような立場のユーザー視点からの発表にも，別の需要があるのではないだろうか……

やってみよう

体験者が統計解析者として平均人なのかは疑問だが…こんな人

Python 3 エンジニア認定基礎試験
Python 3 エンジニア認定データ分析試験
PythonCDA

自己啓発でpython勉強してとってみた

SAS ユーザー総会 2019

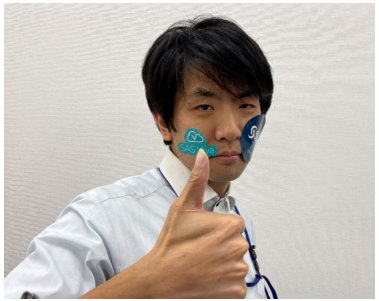
発表者

- 森岡　裕 (35)

  薬業卸で1年営業
  ↓
  CROでデータマネジメント・症例登録センター・統計解析業務を5年
  ↓
  調査会社で，選挙・紙面調査・公官庁統計関連業務を2年
  ↓
  (2016～現在)イーピーエス−統計解析業務

【SAS関連-発表歴】
・SASユーザー総会
→2013～2019で計11本論文発表
・SAS 社会科学研究部会ユーザー会2014発表
・SAS Global Forum 2019

昨年のユーザー総会でのbioに追記



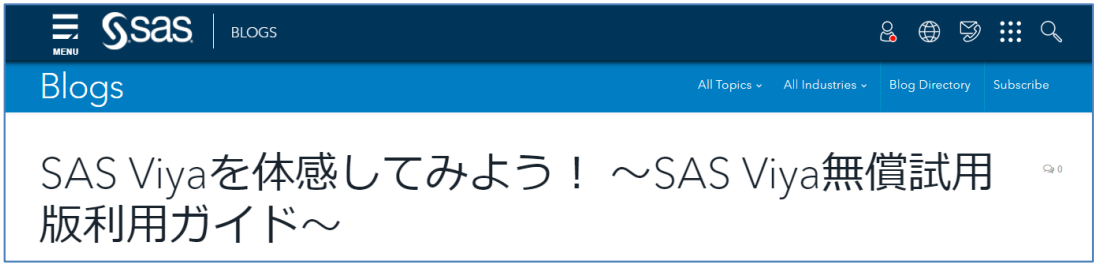
ブログ書いてる

とりあえず，SASプログラミング主体でちょっとpythonかじった程度のスキル

SAS Viya無償試用版について

数年前は14日間の体験期間だったが現在は30日+(自動延長30日)
※ただ，この辺り，利用条件等は，随時更新されているようです

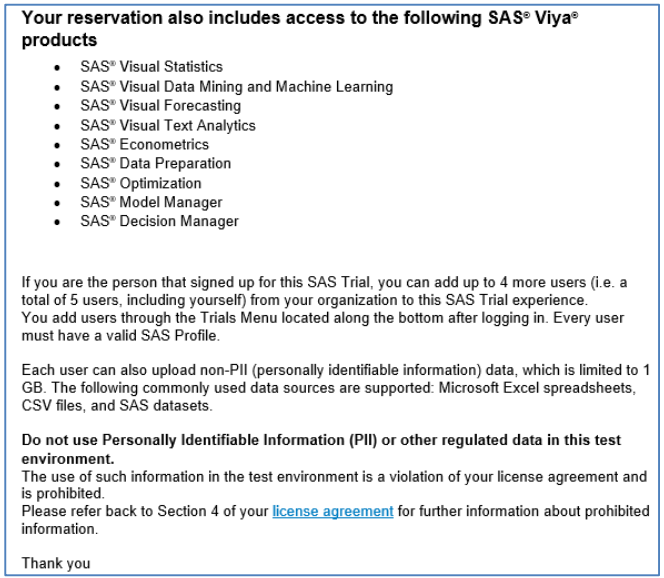
https://blogs.sas.com/content/sasjapan/2017/06/21/sas_viya_free_trial_guide/



感想としてSAS Japan blogでの解説が一番スムーズでした

基本的に連動しているようで，全てを登録する必要はないようでした

チームで体験できる



Your reservation also includes access to the following SAS® Viya® products

- SAS® Visual Statistics
- SAS® Visual Data Mining and Machine Learning
- SAS® Visual Forecasting
- SAS® Visual Text Analytics
- SAS® Econometrics
- SAS® Data Preparation
- SAS® Optimization
- SAS® Model Manager
- SAS® Decision Manager

If you are the person that signed up for this SAS Trial, you can add up to 4 more users (i.e. a total of 5 users, including yourself) from your organization to this SAS Trial experience.
You add users through the Trials Menu located along the bottom after logging in. Every user must have a valid SAS Profile.

Each user can also upload non-PII (personally identifiable information) data, which is limited to 1 GB. The following commonly used data sources are supported: Microsoft Excel spreadsheets, CSV files, and SAS datasets.

Do not use Personally Identifiable Information (PII) or other regulated data in this test environment.
The use of such information in the test environment is a violation of your license agreement and is prohibited.
Please refer back to Section 4 of your license agreement for further information about prohibited information.

Thank you

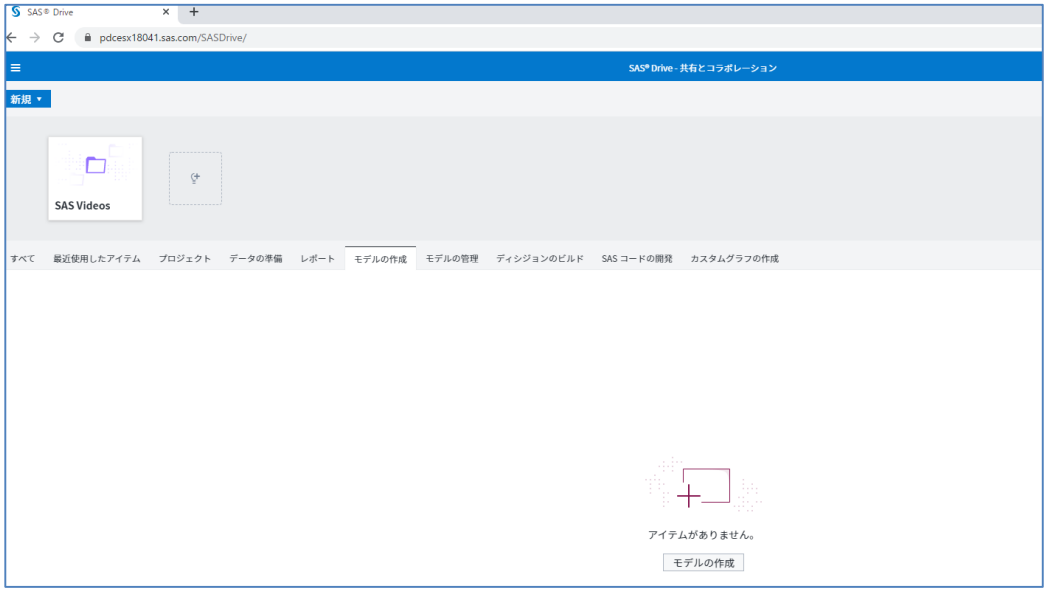
If you are the person that signed up for this SAS Trial, you can add up to 4 more users (i.e. a total of 5 users, including yourself) from your organization to this SAS Trial experience.

5人で共有した環境でviyaを体験できるので，組織に導入を検討する上で非常に有用

Each user can also upload non-PII (personally identifiable information) data, which is limited to 1 GB. The following commonly used data sources are supported: Microsoft Excel spreadsheets, CSV files, and SAS datasets.

ただし，1GBのデータ制限には注意

さて，ログインすると，どうしたものかなという画面になる…

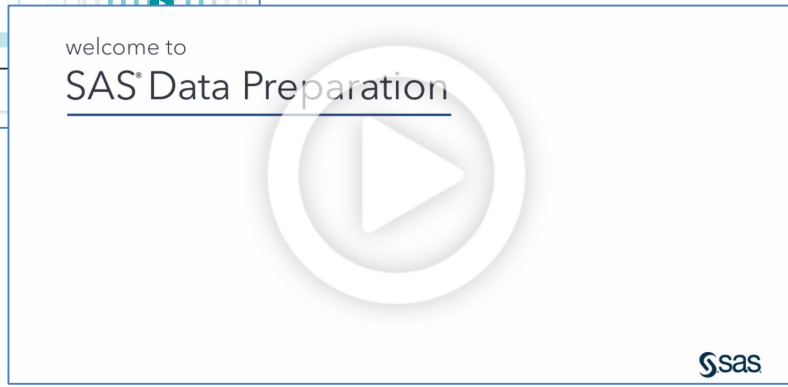


真っ白…

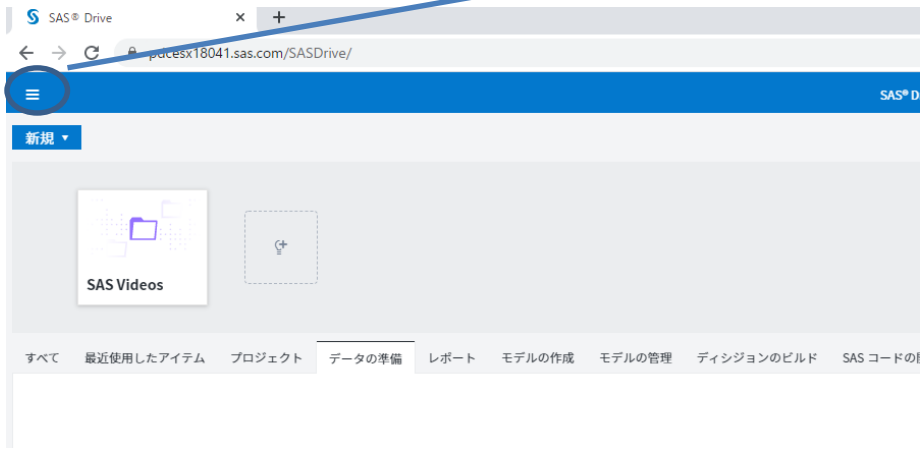
いきなり何をしていいかがわかりにくい…

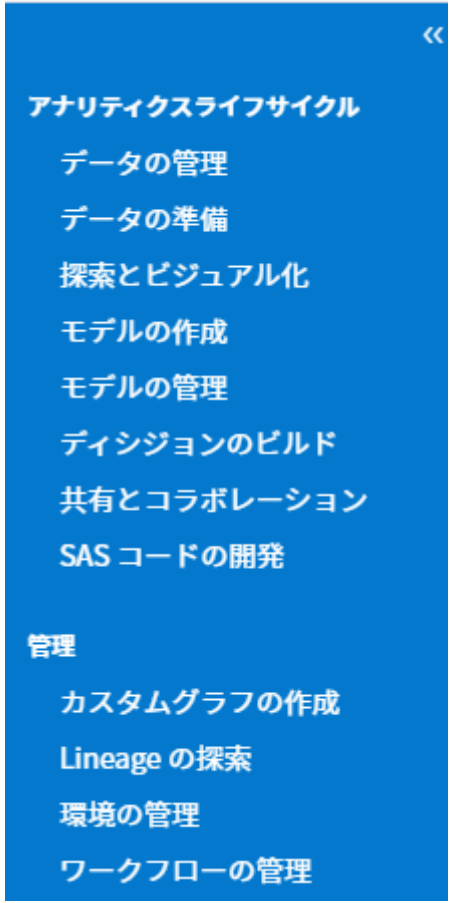
日本SASユーザー会 フォーラム 2020

とりあえず，チュートリアルビデオがあるので見てみる



英語．使い方というよりかは機能紹介に近い？
ちょっとわかりにくい
スタートアップガイドやチュートリアルがもう少し充実しているといいなぁ…

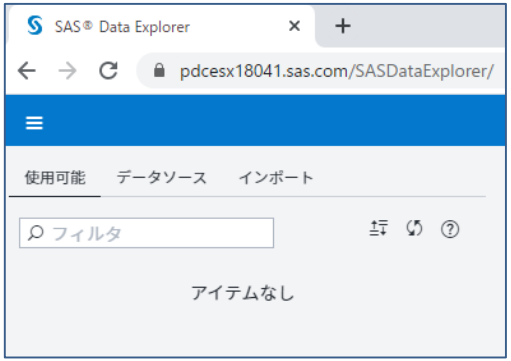




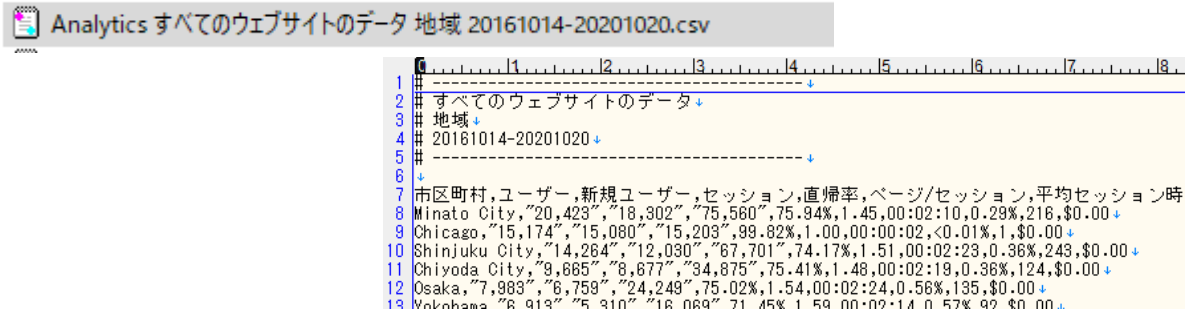
左上の「三」←すごくよく押す

なんとなく，この順番に従って，やってけばいいのかということがわかる

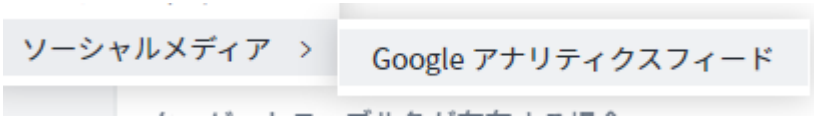
データの管理



インポート→ローカルファイルで予め用意しておいた某ブログのアクセスデータのcsvを指定

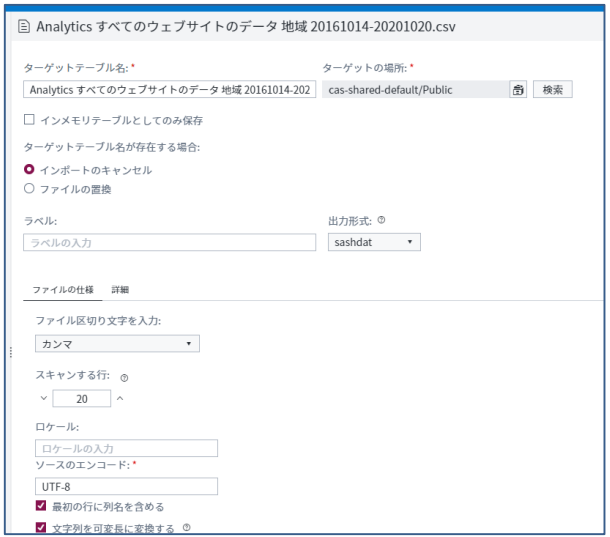


製品版だと，ソーシャルメディアからダイレクトリンクできて，Google Analyticsともリンクしている模様

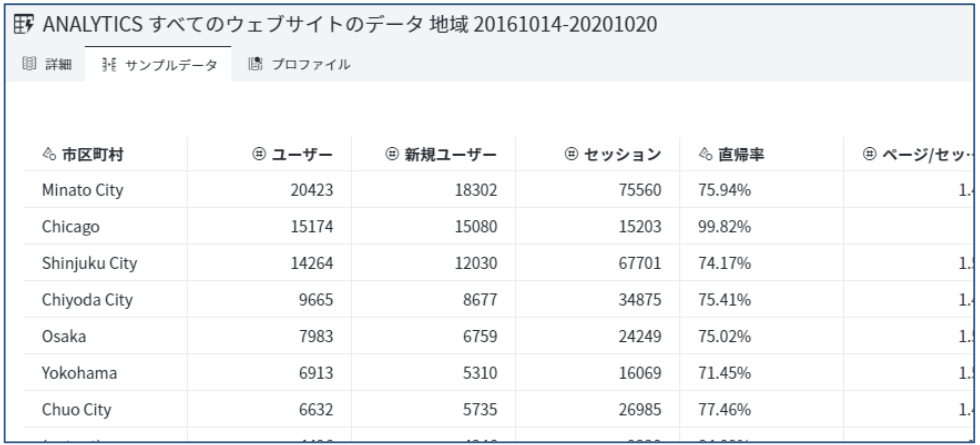


データの管理

SASBASE(SAS ACCESS)のインポートウィザードに近い
(読込開始行オプションが欲しかった…)



#	名前	ラベル	種類	元の長さ	フォーマッ…	出力形式
1	市区町村		varchar	26	26	
2	ユーザー		double	8	12	
3	新規ユーザー		double	8	12	
4	セッション		double	8	12	
5	直帰率		varchar	7	7	
6	ページ/セッション		double	8	12	
7	平均セッション時間		varchar	9	9	
8	コンバージョン率		varchar	6	6	
9	目標の完了数		double	8	12	
10	目標値		varchar	5	5	

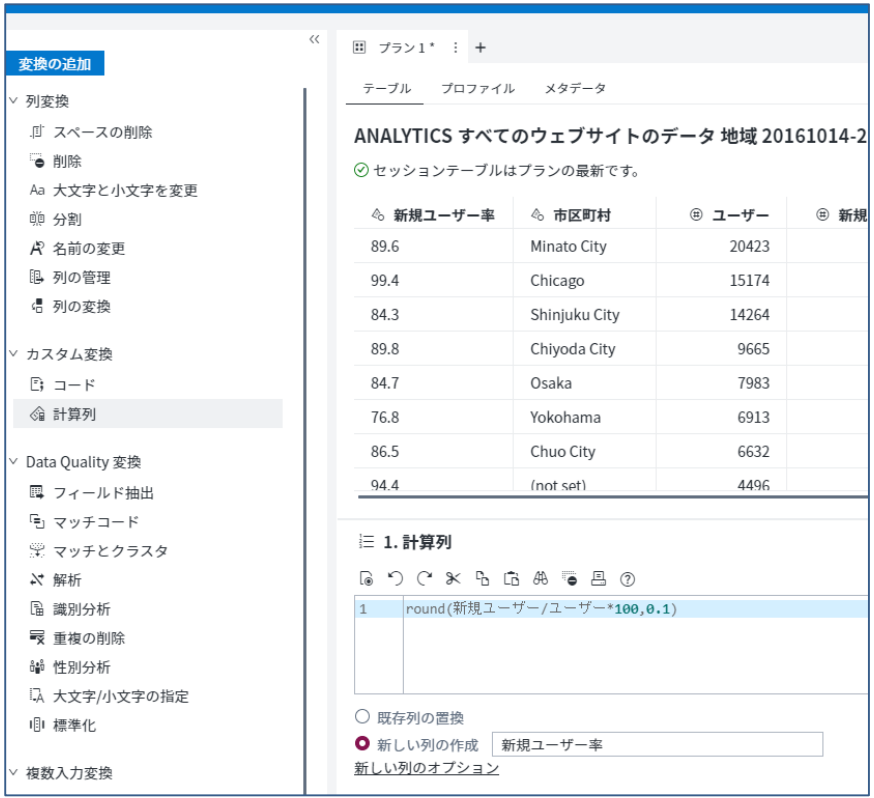


市区町村	ユーザー	新規ユーザー	セッション	直帰率	ページ/セッ…
Minato City	20423	18302	75560	75.94%	1.
Chicago	15174	15080	15203	99.82%	
Shinjuku City	14264	12030	67701	74.17%	1.
Chiyoda City	9665	8677	34875	75.41%	1.
Osaka	7983	6759	24249	75.02%	1.
Yokohama	6913	5310	16069	71.45%	1.
Chuo City	6632	5735	26985	77.46%	1.

データの準備

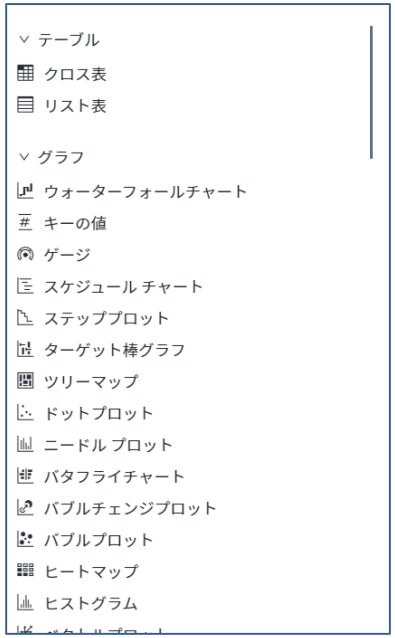
「データの管理」でつくったデータを「データの準備」で整える.
データハンドリングについてもGUIをフルに活用できる

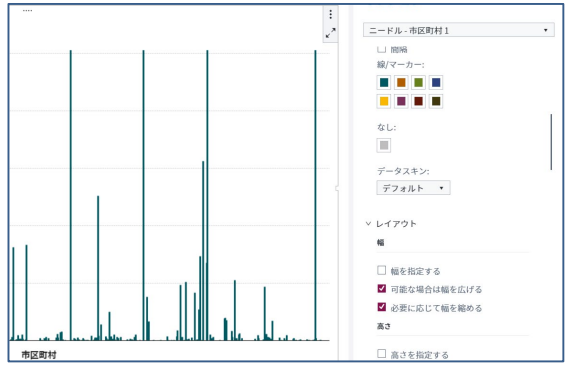
データの準備



GUIでの操作のバリエーションが多く，段階的に他種のハンドリングが実装できる

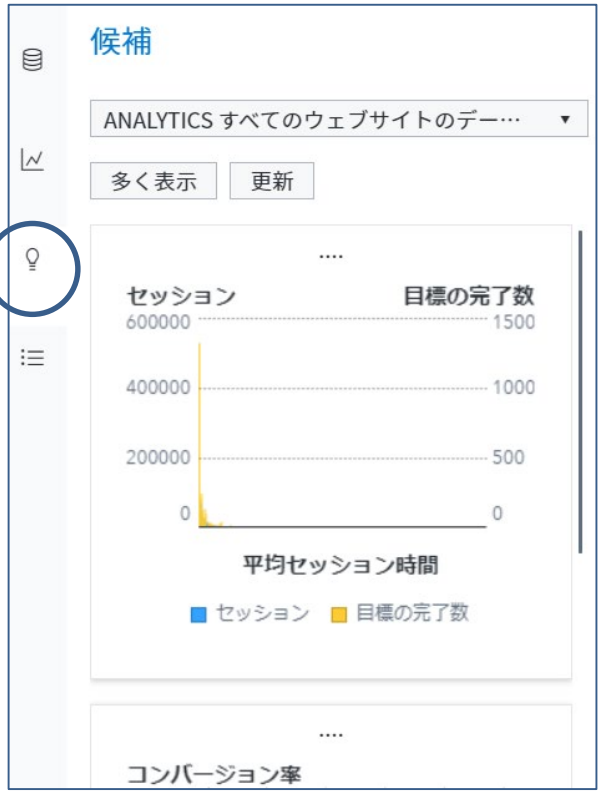
探索とビジュアル化



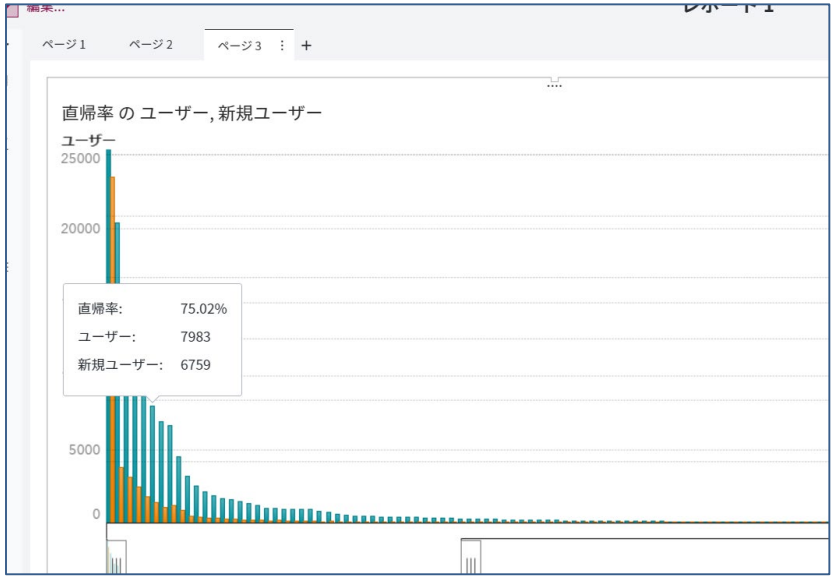


データ，または描きたいプロットから選択して，変数をあてはめ，オプションを設定していくだけで，簡単にビジュアル化ができる

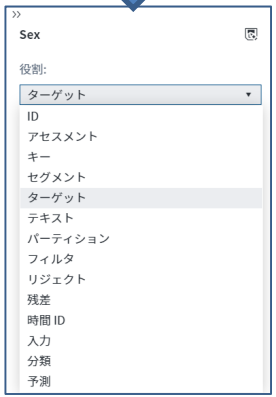
探索とビジュアル化

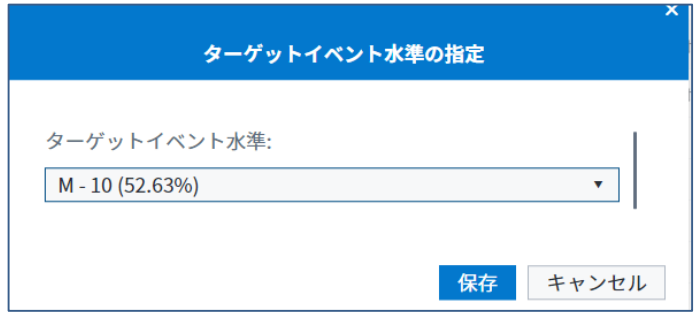


個人的に興味深かったのが「候補」の機能，データを指定しているだけなのに，こういう変数使って，こんなグラフどうですか？という提案が結構，的を得てて驚いた



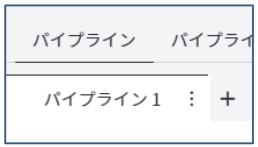
モデルの作成

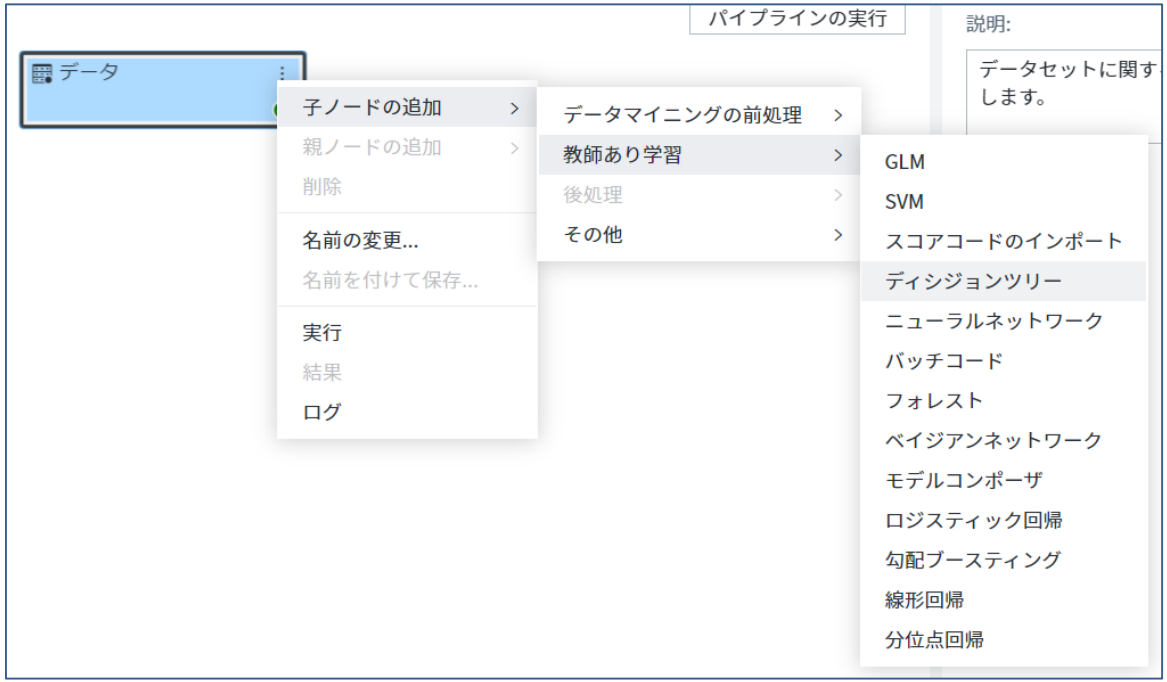




M(男子)を予測するモデル

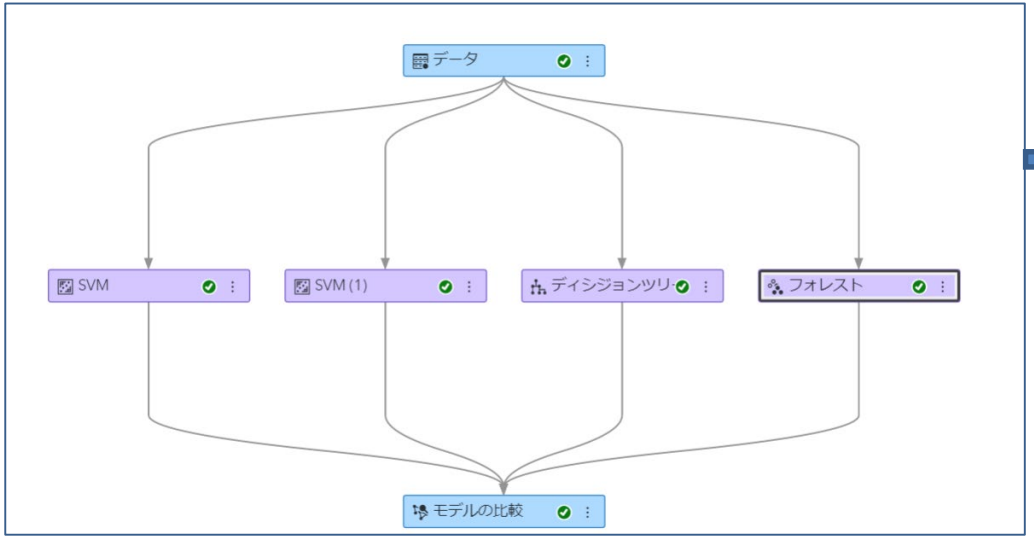
モデルの作成





データを軸として子ノードの追加で簡単に，前処理や学習を紐づけることができる

モデルの作成



特になんの設定をいじらくても，さくっと進んで
チャンピオンモデルが選択される

モデルの作成

sas_class
データ
パイプライン
パイプラインの比較
インサイト
フィルタ
データ: テスト
チャンピオン
名前
アルゴリズム名
パイプライン名
KS (Youden)
度数
SVM (1)
SVM
パイプライン 1
0
当てはめ統計量
統計量
学習
検証
テスト
正解率 1 0.3333 0.5000
誤差 0 0.6667 0.5000
感度 1 0.6667 1
特異度 1 0 0
学習結果
統計量
説明
nSupport サポートベクター数
nSupportInM マージン上のサポートベクター数
MaximumF 最大 F 1.0000
MinimumF 最小 F -1.0000
nEffects 効果数
nLevels データ行列の列数
MatrixSize カーネル行列の列数 1,711

モデルの作成

パス EP スコアコード

```
   dcl nchar(4) "_WARN_" having label n'警告';
   dcl double EM_EVENTPROBABILITY;
   dcl nchar(1) EM_CLASSIFICATION;
   dcl double EM_PROBABILITY;
   varlist allvars [_all_];


   method init();
      _77HNUG91Z7A6J880VU3Y2ZU2J.setvars(allvars);
      _77HNUG91Z7A6J880VU3Y2ZU2J.setkey(n'D7B5CE493147EA1B30FEF90768DB4EEFCBF40199
   end;

   method post_77HNUG91Z7A6J880VU3Y2ZU2J();
      dcl double _P_;
```

DS2 パッケージコード

```
package MS_79c0c539f2044e6caa6fdf7d547e3cfb_20OCT2020193227690 / overwrite=yes;
   dcl package score _77HNUG91Z7A6J880VU3Y2ZU2J();
   dcl double "_P_" having label n'ディシジョン関数';
   dcl double "P_SexM" having label n'予測: Sex=M';
   dcl double "P_SexF" having label n'予測: Sex=F';
   dcl nchar(32) "I_Sex" having label n'ターゲット: Sex';
   dcl nchar(4) "_WARN_" having label n'警告';
   dcl double EM_EVENTPROBABILITY;
   dcl nchar(1) EM_CLASSIFICATION;
   dcl double EM_PROBABILITY;
   varlist allvars [_all_];


   method init();
      _77HNUG91Z7A6J880VU3Y2ZU2J.setvars(allvars);
```

学習コード

```
*------------------------------------------------------------*;
* Macro Variables for input, output data and files;
  %let dm_datalib =;
  %let dm_lib     = WORK;
  %let dm_folder  = %sysfunc(pathname(work));
*------------------------------------------------------------*;
*------------------------------------------------------------*;
  * Training for svm;
*------------------------------------------------------------*;
*------------------------------------------------------------*;
  * Initializing Variable Macros;
*------------------------------------------------------------*;
%macro dm_unary_input;
%mend dm_unary_input;
%global dm_num_unary_input;
```

プログラミング面からも攻めるなら
DS2の学習はしておいた方がよい

モデルの作成

Pythonで少し機械学習などをいじったレベルからの所感として，GUIは非常に使いやすく，Pythonよりさらに平易にモデリングができるといった感想です．

ただ，正直なところ，「良くも悪くも」簡単といった感想をいだきました．

言ってしまえば，SVMがなにかをまったく知らなくてもフローに埋め込めば，やってくれるわけで，複数のモデルの比較についても自動で行われてチャンピオンモデルを提示してくれるので，逆に危険かもしれない…

モデルの作成

英語のハードルは少しあるかもしれませんが…

オンラインコース「Machine Learning Using SAS Viya」

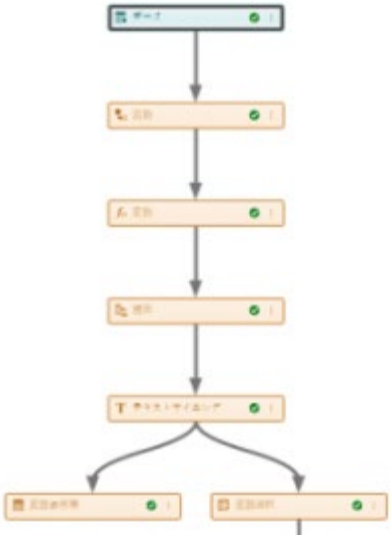
https://www.coursera.org/learn/machine-learning-sas

- Week1：Getting Started with Machine Learning using SAS® Viya®
- Week2：Data Preparation and Algorithm Selection
- Week3：Decision Tree and Ensembles of Trees
- Week4：Neural Networks
- Week5：Support Vector Machine
- Week6：Model Deployment

モデルの作成

Machine Learning Using SAS Viya

なにが，おススメかというと，基本的な理論や考え方，
実務での流れ，そしてViyaで実作業として，どのようなパイプラインを組むのかについて，流れでレクチャーしてくれるので非常に有難いと思います



日本SASユーザー会
フォーラム
2020

AI実用化のための実践ガイド～SAS Viyaで はじめるアナリティクス・ライフサイクル入門～

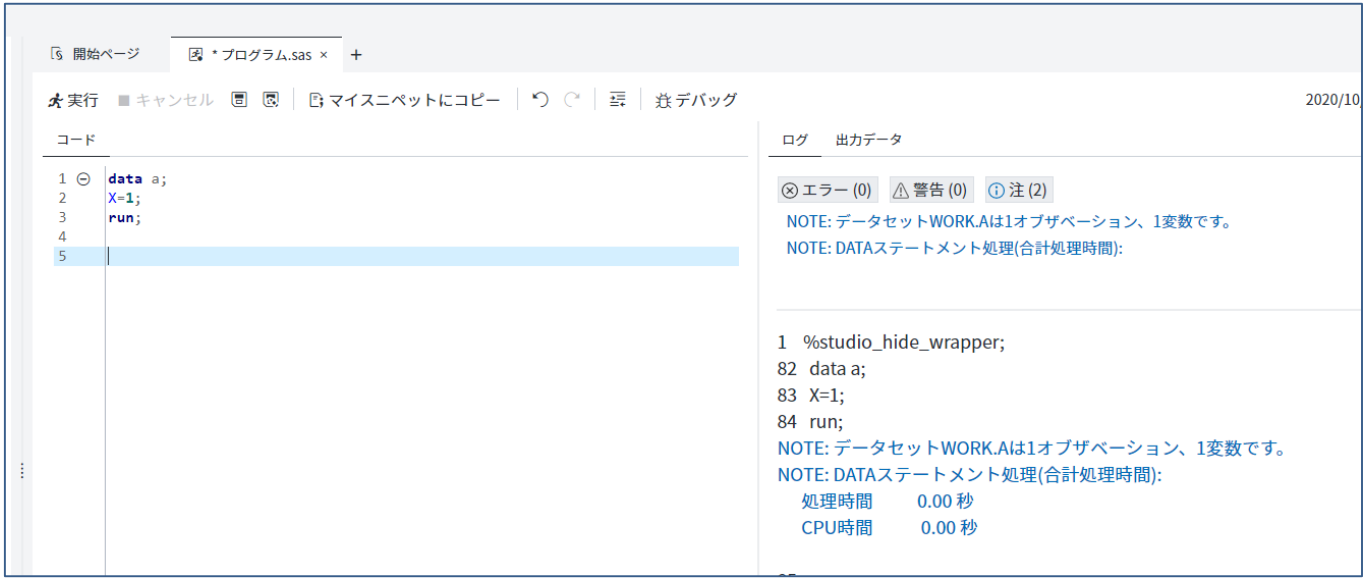
■目次

第1章　アナリティクス・ライフサイクルとは？
第2章　アナリティクスのためのデータマネジメント
第3章　ビジュアライゼーションによるアナリティクスの加速
第4章　機械学習の概要と予測モデリング
第5章　ビジネスで使える自然言語処理
第6章　時系列予測と最適化
第7章　スコアリングとモデル管理
第8章　AI/IoT時代におけるリアルタイム意思決定

Viyaのフローと業務のフローを同期して書かれているので
非常にわかりやすい（ビジネス分野寄り）

SASコードの開発

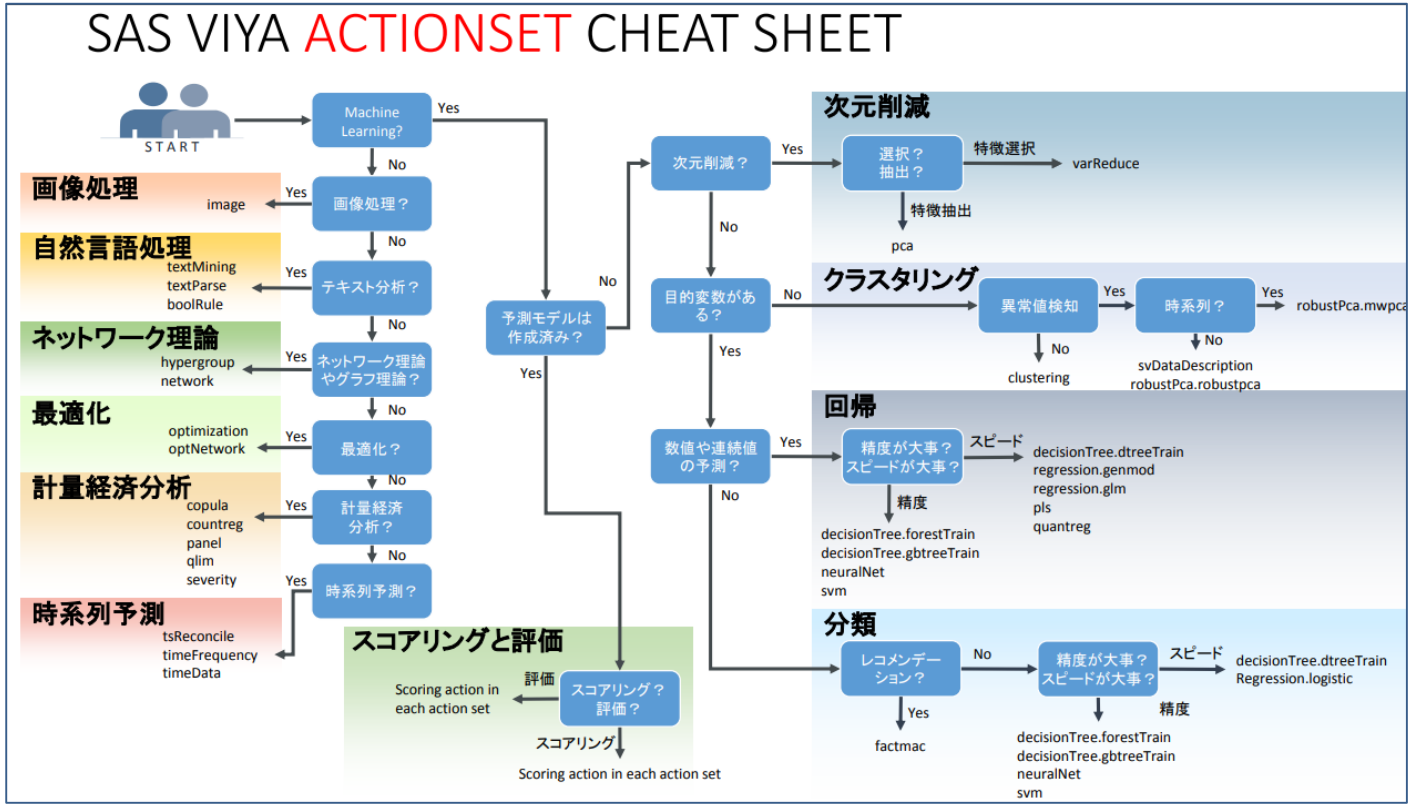
アナリティクスライフサイクル
- データの管理
- データの準備
- 探索とビジュアル化
- モデルの作成
- モデルの管理
- ディシジョンのビルド
- 共有とコラボレーション
- SAS コードの開発



いやぁ…，落ち着く…(笑) まんまSAS Studioですね．ここは通常のSASと同様です
基本的にGUIで可能なことはプログラミングでも同じことができます．（画像処理とか
ちょっと違うみたいですが）

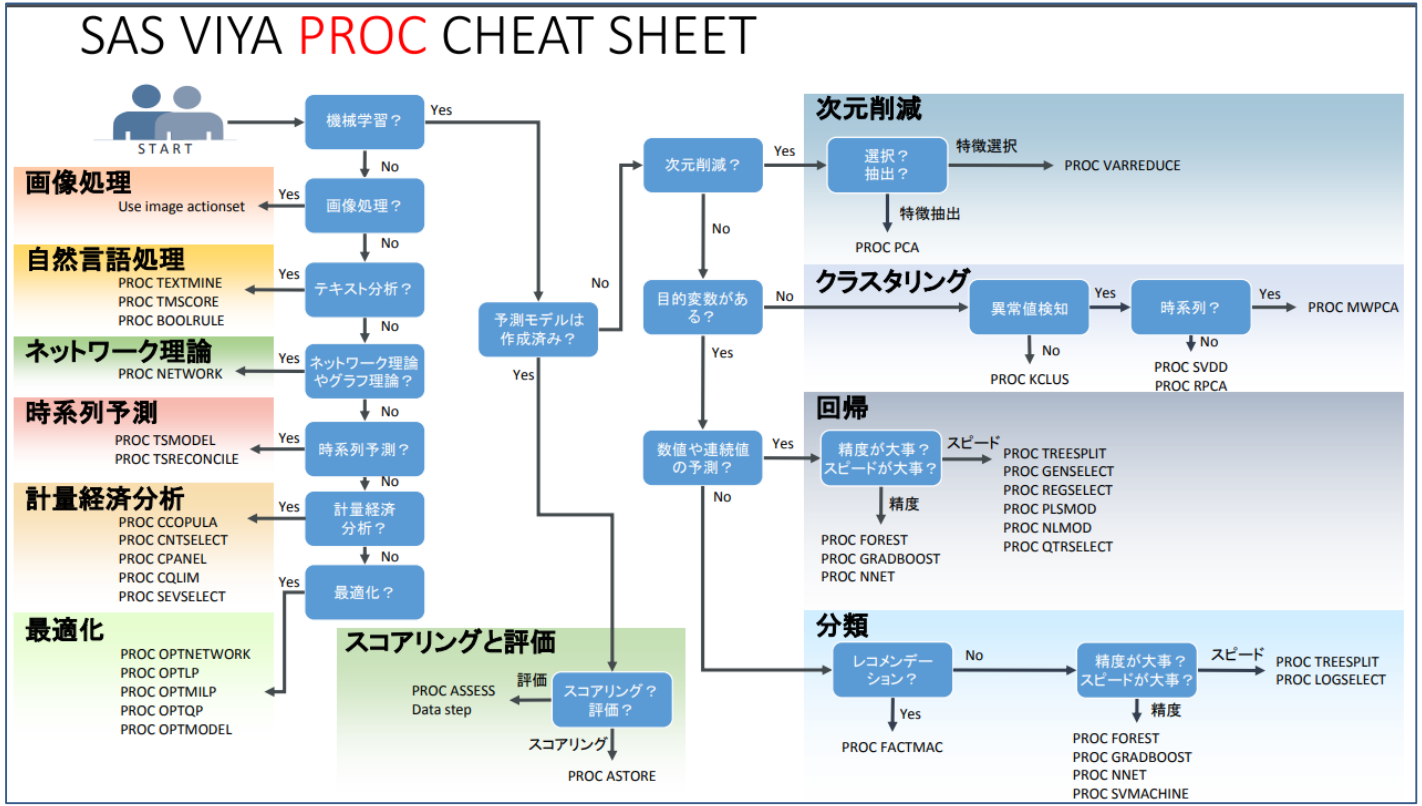
SASコードの開発

https://blogs.sas.com/content/sasjapan/files/2017/07/viya_cheat_sheet_20170721_jp.pdf



SASコードの開発

https://blogs.sas.com/content/sasjapan/files/2017/07/viya_cheat_sheet_20170721_jp.pdf



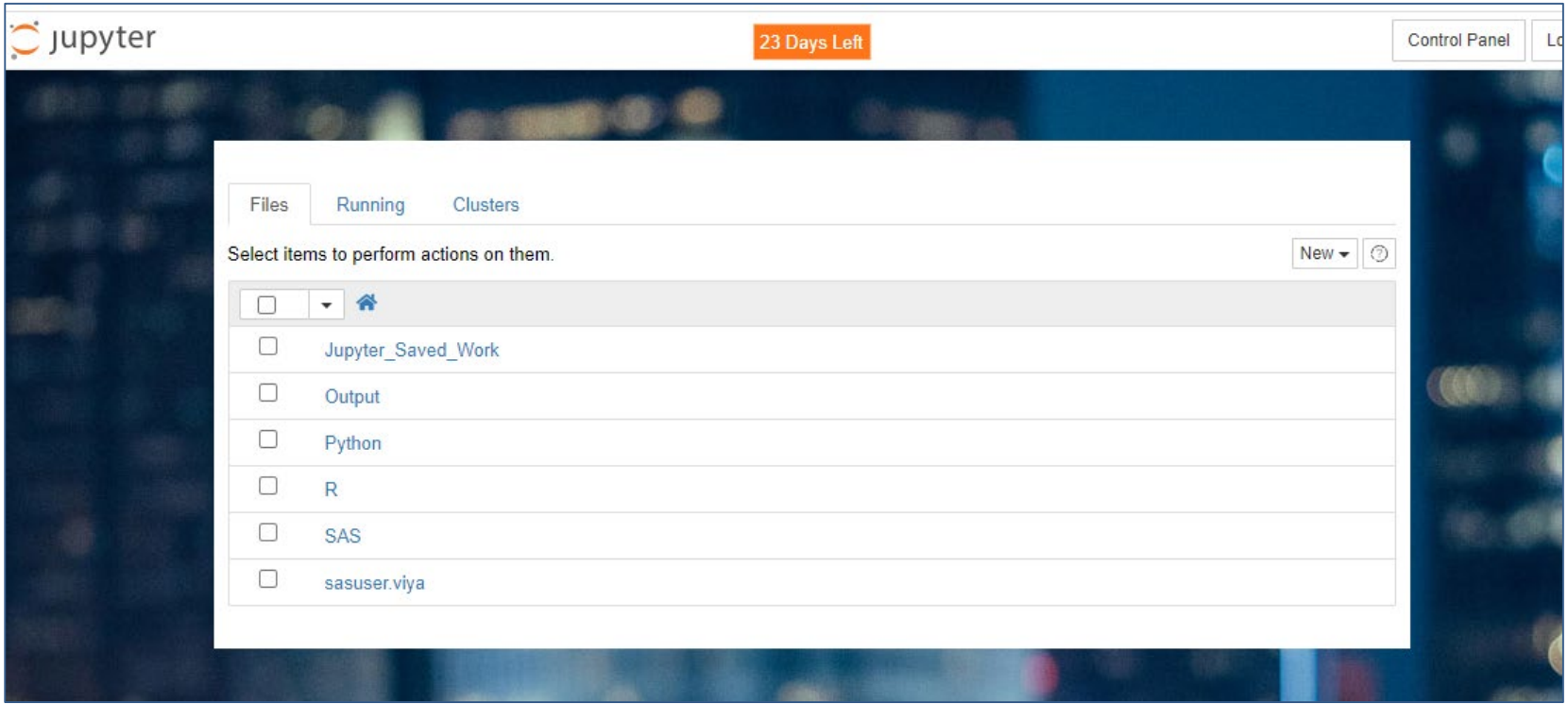
SASコードの開発

一気に使えるプロシジャが増えただけに，目当てのプロシジャを探すのが一仕事でしたが，チートシートのおかげですっきりしました

GUIでの操作と，プロシジャでの指定が同期しているところがまたよいですね

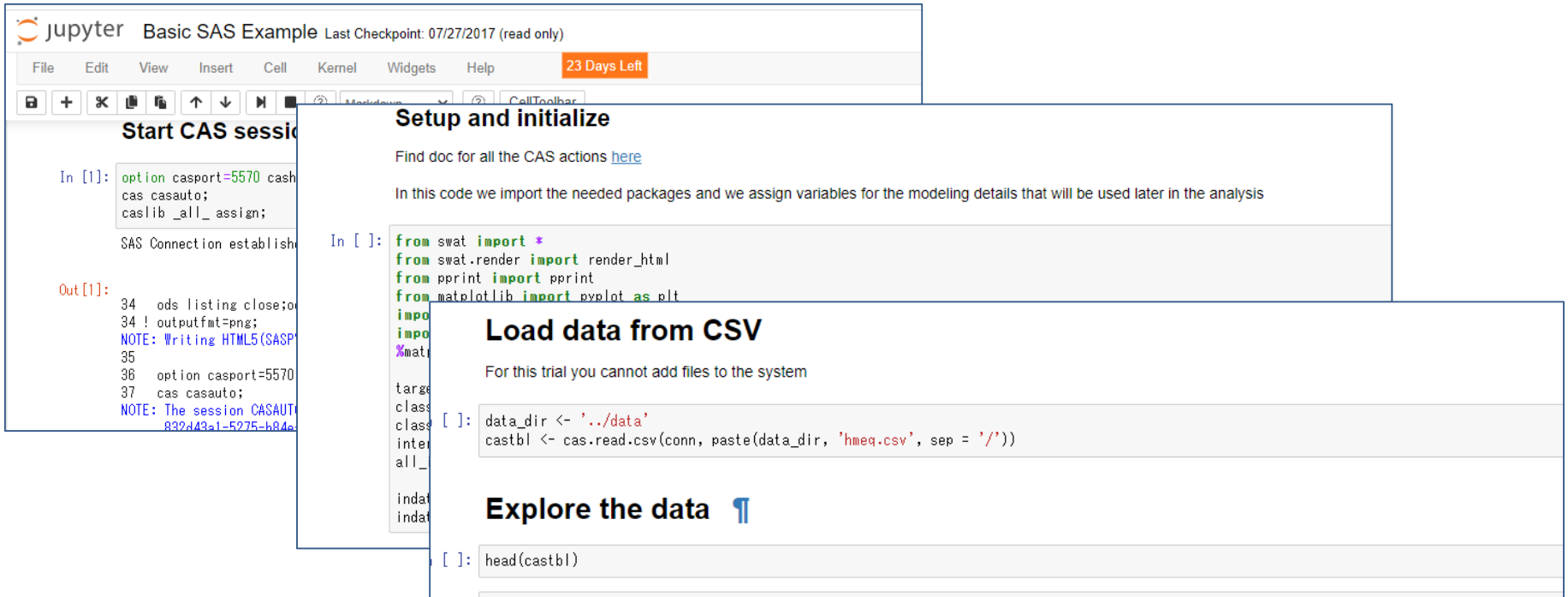
プログラム主体の人もそうでない人もViyaで一緒に作業できるというのも魅力なのかもしれません

SASコードの開発


jupyter
23 Days Left
Control Panel
Files
Running
Clusters
Select items to perform actions on them.
New
Jupyter_Saved_Work
Output
Python
R
SAS
sasuser.viya

SASコードの開発

Jupyter NotebookをインタフェースとしてSAS/R/Pythonでプログラミングができる



まとめ

実は，まだまだViyaでできることを全然説明できておらず，紹介した箇所もさわりだけです．
今回，体験版でいろいろと操作してみて
「オープンなAIプラットフォーム」という看板に偽りなしと感じました．

様々な種類のデータ・様々な統計手法，機械学習，深層学習が扱えるというだけなく
多様なスキルセットをもった人材が共通したプラットフォームとして利用できるというところ
に魅力があると思います．

今後，よりデータは大規模化していきます．またリアルタイムに変化していくことが常になっていくのだと思います．そういった状況に対応することが求められています．さらに，ミクロな視点とマクロな視点を併せ持つこと，AIによるアウトプットに対して，どのように解釈・判断していくかなど，Analyticsの世界はは大きな転換期を迎えていると思います

そういった新しい世界に適したプラットフォームではないかと可能性を感じました

日本SASユーザー会
フォーラム
2020

ご清聴ありがとうございました